JOURNAL OF APPLIED PHYSICS

Physics Library
QC
1
.J863

JOURNAL
OF
APPLIED PHYSICS

Vol. 20, No. 1 JANUARY, 1949

In This Issue

Page

Structure of Extremely Thin Layers Evaporated in Kinetic Vacuum Systems H. A. Stahl 1

Thermal Behavior of Evaporated Layer Vacuum Devices H. A. Stahl 8

Stress-Strain Rate Relations for Anisotropic Plastic Flow John E. Dorn 15

Theoretical Resolution Limit of the Electron Microscope O. Scherzer 20

Internal Friction of Zinc Single Crystals Charles A. Wert 29

Demodulation of an F-M Carrier and Random Noise by a Limiter and Discriminator Nelson M. Blachman 38

Determination of the Elastic Constants of Solids by Ultrasonic Methods William C. Schneider and Charles J. Burton 48

Effect of Current Magnitude upon the Behavior of a Superconducting Bolometer in Its Transition Region Nelson Fuson 59

Effect of Uplift and Denudation on Underground Temperatures A. E. Benfield 66

Experimental Investigation of Formulas for the Prediction of Horn Radiator Patterns G. A. Woonton, D. R. Hay, and E. L. Vogan 71

Incremental Friction Coefficient—a Non-Hydrodynamic Component of Boundary Lubrication J. T. Burwell and C. D. Strang 79

Optical Equations of Three-Dimensional Photoelasticity Raymond D. Mindlin and Lawrence E. Goodman 89

Method for Measuring the Total Power of Small-Angle X-Ray Scattering B. E. Warren 96

Electron-Micrographic Structure of Shadow-Cast Films and Surfaces Robley C. Williams and Robert C. Backus 98

Determination of Stresses in Plastic Regions in Problems of Plane Flow P. S. Symonds 107

Erratum: Transmission Properties of Materials in the Millimeter Wave Region E. Wantuch, J. H. Rohrbaugh, J. P. Cooley, J. H. Greig, and J. Shmoys 112

Heat Transfer in Sweat-Cooled Porous Metals Sidney Weinbaum and H. L. Wheeler, Jr. 113

Letters to the Editor:

New Method of Capacity Measurement on Dry Disk Rectifiers K. Lehovec 123

"Preliminary Study of a Physical Basis of Bird Navigation" William C. Schumacher 123

Method for Studying the Forces between Metals and Ionic Substances Evelyn C. Marboe and W. A. Weyl 124

Effect of "Multiple Grounds" on Electron Microscope Images W. L. Grube 125

In Every Issue

Calendar of Meetings 37

New Books 126

New Booklets 126

Here and There 127

BOARD OF EDITORS
Elmer Hutchisson,
Editor

Associate Editors
R. Bowling Barnes
Otto Beeck
W. James Lyons
Philip Morrison
Paul E. Sabine
Raymond J. Seeger
Frederick Seitz
Charles S. Smith
J. R. Townsend

Publication of the
American Institute of Physics

OFFICERS
George R. Harrison,
Chairman
Wallace Waterfall,
Secretary
George B. Pegram,
Treasurer
John T. Tate,
Adviser on Publications

GOVERNING BOARD
Luis W. Alvarez
Robert B. Brode
E. U. Condon
A. J. Dempster
W. F. Fair, Jr.
Harvey Fletcher
R. C. Gibbs
George R. Harrison
W. V. Houston
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
George B. Pegram
A. H. Pfund
John T. Tate
Wallace Waterfall
M. W. White
A. G. Worthing

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton,
Director
Cleveland Norcross,
General Manager
Theodore Vorburger,
Business Manager
Ruth F. Bryans,
Publication Manager
Betty Hopstadter,
Office Manager
Eileen B. Neuberger,
Circulation Manager

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$7.00	$ 8.00
To all others	9.00	10.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 19, 1948—$140.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879.
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1949, by the American Institute of Physics.

Physics Lib.
Falcon

Journal
of
Applied Physics

Volume 20, Number 1 January, 1949

The Structure of Extremely Thin Layers Evaporated in Kinetic Vacuum Systems*

H. A. STAHL**
Laboratory for Material Physics, Telefunken Tube Factory, Berlin, Germany
(Received March 28, 1947)

In a commercial electron diffraction camera some alkaline earth metals (from Be to Sr) or Al, Ni, or Mo were evaporated onto outgassed metal sheets or glass slides, using the starting pressure of 0.1μ of mercury. The diffraction patterns of the thinnest layers (to about 10^3A) presented, regardless of the evaporation speed used, the pattern of the oxide of the evaporated metal exclusively. The diffraction pattern of the metal space lattice itself was growingly observed with Be, Al, Ni, and Mo only after *uninterrupted* evaporation until layer thicknesses of about 1μ were reached. When Mg, Ca, and Sr were evaporated, no patterns without, at least, the most intensive oxide rings could be attained under the vacuum conditions given in the apparatus.

The results are in accord with some earlier findings of Langmuir, Beeching, Darbyshire, and Beischer, and clear up some hitherto not fully explained effects with stepwise evaporated layers. The appearance of oxide patterns is always accompanied by strong improvements of the vacuum in the evaporation compartment by the arising metal vapors, i.e., by the well-known gettering effect.

I. PROBLEM

IN the large-scale production of electronic tubes of all kinds, many metals (as, e.g., nickel, iron, molybdenum, tungsten, or the alkaline earth metals) are heated to high temperatures in vacuum systems. The metal evaporates according to its vapor pressure curve and condenses again from the gas phase. The evaporation may occur either as an undesired secondary effect (sputtering of incandescent tungsten filaments in lamps or evaporation of overheated grids) or as a desired one (preparation of getter layers or the production of mirrors for optical purposes).

In every case, these processes and their consequences are of great importance in the production process itself, as well as in the later operation of the tube; their scrutiny therefore becomes a technical necessity. Since very thin layers are produced by evaporation, electron diffraction is a favorite tool to obtain qualitative, as well as quantitative, data about the structure of these films and, moreover, about processes comprising the evaporation itself.

The pressure in the tube is of the greatest importance. It may be measured directly so long as the tube is attached to the vacuum manifold; the pressure of the sealed-off tube may be evaluated from its vacuum factor. Direct measurements on pumping machines gave pressures up to 10^{-2} mm Hg; from vacuum factor measure-

* This document is based on work performed 1941 to 1943 within the research program of a foreign worker physicist in Germany. The printing was delayed by the post-war conditions in Europe.

** Now at the US-Army Technical Information and Education School, Stuttgart-Vaihingen, Germany. Ex-member of the Czech Association of Physicists and Mathematicists, and of the Czechoslowak Testing Corporation, both Prague, and of the German Society for Technical Physics, Berlin.

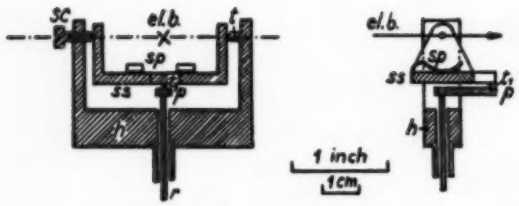


FIG. 1. Specimen holder and tilting design. The seesaw, *ss*, is supported between the screw, *sc*, and the tip, *t*, the U-shaped fork, *h*, being liftable and sinkable as a whole. Tilting ensues by a side projection, *p*, and a tip, t_1, separately movable by the rod, *r*. The experimental arrangements are held by springs, *sp*.

ments on tubes of various types G. Herrmann and O. Krieg[1] estimated the pressure in sealed-off tubes as about 10^{-5} mm Hg. M. Littmann[2] reports a residual gas pressure from 10^{-5} to 10^{-4} mm, chiefly caused by the evolution of decomposition gases from the softened seal-off constriction. In producing tubes on a large scale, the latter are connected to the pumps by means of rubber hoses which cannot be regarded as unobjectionable even if they contain cavities to be pumped up.

The vacuum beneath bell jars, often used for experimental work, cannot be regarded as satisfactory at all. Neither the jar, nor the base plate is outgassable, and if a rubber gasket is present, the very large circumference entails considerable amounts of air constantly entering the evacuated space. To this class of vacuum devices belong commercial evaporating chambers, electron microscopes and diffraction apparatus, vacuum spectrographs, and x-ray equipment.

If the tube is blown to the high vacuum pump, or if it is sealed off, a static vacuum system is obtained; in all the above-mentioned cases, however, kinetic vacuum systems are under consideration, the vacuum of which can only be maintained by the continuous operation of a high speed pump, especially in the case of demountable apparatus with many sections, bushings, motion devices, etc. It must be emphasized, moreover, that such vacuum devices are never outgassable. Results obtained with such systems must not be supposed to hold in the complete absence of gas. Nevertheless, such arrangements are often designated as "high vacuum devices" in the literature, though rather incorrectly, since, for example, W. Paul,[3] upon exciting the residual gas of $10^{-2}\mu$-Hg by electron impact, observed a distinct emission which gradually disappeared only below that pressure.

The present electron diffraction investigation was carried out in a commercial electron diffraction apparatus at about the same pressure range as that obtained in mass-produced vacuum tubes which are still attached to the pumps. Therefore, the results obtained will be of technical importance in vacuum tube manufacture.

The investigation was initiated by the phenomenon, exhibited by numerous diffraction patterns, that all the interior surfaces of vacuum tubes in which small quantities of magnesium getter had been evaporated showed exclusively the pattern of the MgO space lattice. The question at once arose as to whether or not the surfaces of the intact, exhausted tube are already covered with magnesium oxide, or whether the oxide is formed when the bulb is opened to the atmosphere. Here it may be mentioned preliminarily that Mg films above an inferior threshold thickness always gave the pattern of the metal lattice, even when the film had been stored in the open air for a long time.

II. EXPERIMENTAL ARRANGEMENT

The electron diffraction apparatus of H. Seemann[4] was used, exhausted by a three-stage

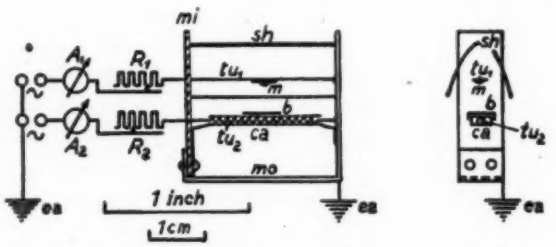


FIG. 2. Design for coating outgassed bases, *b*, by minute quantities of condensed metal vapors. The device is held by a molybdenum bridge *mo* and a mica wing *mi*. The metal sample *m* to be evaporated is welded to the tungsten filament, tu_1, and surrounded by a sheet metal shelter, *sh*, to protect the inner surface of the specimen chamber from being sedimented by metal. The intercepting plane, *b*, is fixed onto a nickel casing, *ca* (5×20×2 mm), and heated by an alumina-coated tungsten coil, tu_2. The arrangement is earthed.

[1] G. Herrmann and O. Krieg, Die Telefunkenröhre, No. 21/22, 219 (1941).
[2] M. Littmann, *Getterstoffe und ihre Anwendung in der Hochvakuumtechnik* (C. F. Wintersche Verlagsbuchhandlung, Leipzig, 1938), p. 13.
[3] W. Paul, Zeits. f. Physik **117**, 775 (1941).
[4] H. Seemann, Zeits. f. Physik **92**, 270 (1934).

Leybold mercury diffusion pump. The apparatus, constructed of many parts sealed by means of rubber gaskets, was maintained under a kinetic vacuum of 10^{-4} mm Hg, but was not, of course, outgassable.

The acceleration voltage was 80 kv, the corresponding de Broglie wave-length, 0.0417A, and the surface layer contributing appreciable intensities to the pattern 10^{-2} to 10^{-3} mm thick. A Greinacher Stabilivolt plant, supplied with 500-cycle alternating current and completed by an electronic regulator, guaranteed a voltage constancy of better than 0.1 percent during the exposure. The voltage was measured by a moving coil milliammeter through a high ohmic resistance, and was often standardized by photographic observation of the diffraction pattern of gold foils.

The object holder could be moved in a plane perpendicular to the electron beam. In the calculation of the lattice constant of the surface material investigated, the camera length is involved. The original tilting mechanism of the Seemann apparatus (necessary to introduce the specimen surface into the beam at the correct angle) incorporated a lever arm of almost 15 cm in length. The focusing of the specimen surface was accompanied, therefore, by an undesired variation of camera length, amounting to 5.2 mm/degree. Another specimen holder was constructed, allowing rotation of the specimen surface about an axis perpendicular to the beam and lying in the surface itself (see Fig. 1).

The experimental arrangement, at first used exclusively with magnesium, is shown in Fig. 2. The distance from the metal sample, *m*, to be evaporated to the base sheet, *b*, on which the metal was to be deposited was uniformly about 1 cm.

The masses to be evaporated were so minute that tungsten coils became unsuitable, because the small metal samples often fell out of them. The metal was welded on a small nickel sheet (1 by 1 mm) using hydrogen atmosphere to prevent oxidation. The nickel sheet was then welded to the tungsten filament, tu_1.

The metal vapors were condensed upon previously outgassed nickel, iron, or platinum sheet metal bases, *b*, but sometimes also on mica or glass slips. The suitably shaped sheet metals were

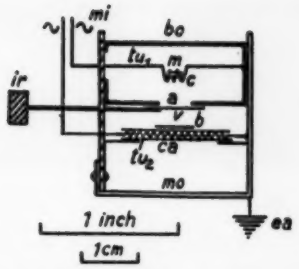


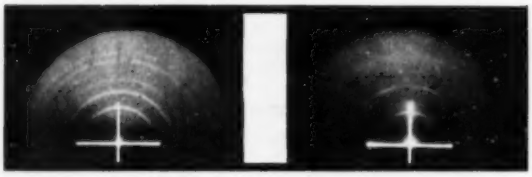
FIG. 3. Design for evaporation of larger quantities of metal, preventing the interception plane, *b*, from being coated with the first evaporation rate. The metal sample, *m*, inside a tungsten coil, *c*, is surrounded by a nickel box, *bo*, the aperture, *a*, of which can be magnetically opened by an iron block, *ir*. Otherwise the letters mean the same as in the preceding figure.

laid on nickel casings, *ca*, as used in indirectly heated vacuum tubes, and heated by alumina-insulated tungsten coils tu_2. If mica or glass was used as the intercepting plane, *b*, the latter could not be thoroughly outgassed when the temperature was kept below that of deterioration of the material. Results obtained from evaporation experiments to these various base sheet materials did not differ perceptibly from each other.

In order to prevent the highly polished interior walls of the specimen chamber from being coated with evaporated metal, the filament, or heating coil, was surrounded by a semicylindrical or parabolic shield, *sh*, and later by an entirely closed casing box, *bo*.

The design shown in Fig. 3 was used when larger masses of metal had to be evaporated and the initial evaporation, with possible contaminating traces, could not be allowed to reach the intercepting sheet, *b*. In this case the slide valve, *v*, about 3 mm distant from the base being used, blocked the aperture, *a*, of the box and was not removed until the evaporation had proceeded for some time ($\frac{1}{2}$ to 1 min.). The iron block, *ir*, was magnetically attracted toward the glass window of the object chamber to effect the removal of the shutter from the aperture.

Agfa Electronic or Agfa Blue-spectral-rapid plates were used. The photographs were measured by means of a Zeiss glass scale, divided into half-millimeters. The center of the diffraction rings was marked on the plate by means of a coordinate cross, drawn by magnetic deviation of the beam into two directions perpendicular to each other. Of course, the specimen was pre-



MgO (a) Mg (b)

FIG. 4. Diffraction patterns of layers, produced by evaporation of about 10^{-4} gram (a) or 10^{-2} gram or more (b) of magnesium.

viously removed from the electron beam. Occasionally a very slight shift of the origin so fixed was observed, relative to the real center of the diffraction rings.

III. RESULTS

All the layers investigated were evaporated in the Seemann apparatus proper and were immediately afterwards investigated by electron diffraction.

Alkaline earth metals from beryllium to barium, inclusive, were evaporated, in addition to typical vacuum tube materials such as molybdenum and nickel. The behavior of aluminum also was of interest in regard to a supposed suppression of some undesired grid emission.[5] All the metals used were at least technically pure, contaminations like Mg, Mn, Ca were, for the most part, only spectroscopically traceable, and Fe and Si were chemically determined as less than 0.2 percent.

Layers obtained by evaporation of not more than 10^{-4} g of magnesium were investigated first and in greatest detail. The diffraction patterns of the deposited material showed a cubic face-centered space lattice of lattice constant 4.20A. This lattice must be ascribed to MgO. In some cases the ring system of the base also appeared, corresponding to a layer thickness of less than 200 unit cells. If the mass of evaporated metal exceeds 10^{-4} g, the most intense diffraction ring of magnesium space lattice begins to emerge Assuming that the entire sample has been evaporated into the half-space, the thickness of the deposited film is computed to be about 10^{-4} mm. Layers from 10^{-4} to 10^{-3} mm, prepared by the arrangement shown in Fig. 3, yield the pattern of the metallic magnesium lattice in greater intensity than that of the oxide. However, under the vacuum conditions, obtaining the production of the metal space lattice pattern alone was never achieved (see Fig. 4).

The same results were obtained with calcium and strontium, likewise used as getter metals. Beryllium and the other metals investigated also yielded only the oxide pattern for the thinnest films; coats thicker than a few microns gave the pattern of the metal space lattice exclusively (see Fig. 5). Coats of intermediate thickness showed both patterns, a gradual transition from the pure oxide to the pure metal lattice seeming to occur.

The thinnest layers obtained by the evaporation of aluminum gave either a diffuse scattering effect or one to two vague rings which could be recognized as belonging to the 004 and 133 interference of γ-Al_2O_3.

Unexpected circumstances seem to occur when Fe-Ba 9 or Fe-Ba 23 getters, both production types, are evaporated. According to the tables of J. D. Hanawalt, H. W. Rinn, and L. K. Frevel,[6] the patterns of $BaSO_3$ and $BaSO_4$ are obtained. Upon heating these coatings to a dull red glow for two minutes, the diffraction ring system of barium sulfide appears as expected.

All the results obtained proved to be independent of the particular base material upon which the metals were evaporated. It may be noted was that magnesium also evaporated upon sprayed or cataphoretically sedimented pure or mixed carbonates (calcium, strontium, or barium) and upon pure or mixed oxide cathodes, evolved from the carbonates by heating them in the apparatus without breaking the vacuum.

IV. DISCUSSION AND FURTHER INVESTIGATION

In view of the results enunciated in the last chapter, we may explain uniformly the behavior of Al, Mo, Ni, as well as the alkaline earth metals (except barium). Extremely thin layers (less than 10^{-4} mm) exhibit diffraction patterns of metal oxide; the ring system of the evaporated metal itself appears only with film thicknesses greater

[5] H. Herrmann, *Der Einfluss von nichtmetallischen Zusätzen auf die Elektronenemission einer Erdalkalioxydkathode* (Dissertation, Polytechnic Academy, Berlin, 1937).

[6] J. D. Hanawalt, H. W. Rinn, and L. K. Frevel, Ind. and Eng. Chem. **10**, 457 (1938).

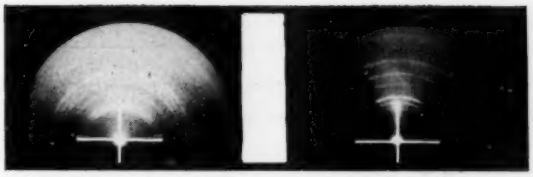
than 10^{-4} mm, the sample being evaporated in a kinetic vacuum system with initial vacuum 10^{-4} mm Hg.

The reasons for this surprising behavior must now be ascertained. The needed oxygen may arise from three different sources:

(a) from the evaporating source,
(b) from the base on which metal is deposited, and
(c) from residual gases in the apparatus.

As to the first point, it may be noticed that the metal to be evaporated and the filament cannot—in contrast to common vacuum practice—be previously outgassed at temperatures higher than those to be used. Under the present experimental conditions, the pressure in the system rose during the first phase of the evaporation (a few seconds) to 10^{-3} mm and then fell rapidly to 10^{-5} mm (see Fig. 6). Also contradicting hypothesis (a) are the results obtained with the apparatus which incorporates the magnetically operated slide valve. If the slide valve is closed until the evaporation proceeds at the lower pressure (after the initial evolution of gas) and then opened to overcoat the base with a thin film, the diffraction pattern shows only the diagram of magnesium oxide. Previous comparable experiments with the aperture open during the entire evaporation, yielded, on the contrary, the ring system of the metal space lattice, accompanied by only the most intense interferences (002 and 022) of MgO. This result forces us to conclude that the oxygen does not issue from the filament and the heated sample, but from the space between the atom source and the intercepting sheet. Upon conclusion of the evaporation, the transparent and brittle skin of the metal sample was left in the coil. An x-ray examination showed the substance to be MgO. Since the oxides of all the metals investigated have much lower vapor pressures than the metals themselves, the chance that any trace of the residual oxide reached the intercepting plane is negligible.

The second hypothesis is disproved by virtue of the careful outgassings of the sheet bases. Accordingly, it is improbable that the impinging metal atoms react with adsorbed or bound oxygen on the base. Furthermore, thick evaporated layers (10^{-3} mm or more) were also used as bases. Such layers exhibited the pattern of the metal



BeO (a) Be (b)

FIG. 5. Diffraction patterns of layers produced by evaporation of about 10^{-4} gram (a) of 10^{-2} gram or more (b) of beryllium.

space lattice (in the case of Mg with slight traces of the most intense 002 and 022 interferences of MgO). An additional thin layer evaporated some minutes afterwards gave only the pattern of metal oxide. Thus, the evaporation of a very small quantity of metal results in the metal oxide pattern, whether metal had previously been evaporated or not.

The oxygen appearing in the evaporated layers as oxide must therefore arise from the residual gases which are continuously present in the kinetic vacuum system.

A thermocouple (Pt-PtRh) welded on the base sheet metal exhibited temperatures of 70 to 100°C during evaporation. This temperature is not high enough for subsequent oxidation of the condensed metal film, as will be shown in the following paper. Hence, the oxidation can occur only during the evaporation process itself.

This result, that the oxygen necessary for forming the oxide film comes from the residual gas in the apparatus, is rather surprising when the number of oxygen atoms present in the apparatus is compared with the amount of oxygen bound in the layer. Taking a gauged pressure of 10^{-4} mm as 2×10^{-5} mm of oxygen, the Seemann apparatus of nearly 3 liters' volume contains 4.3×10^{14} oxygen atoms. On the other hand, the evaporation from the metal vapor sources occurs in the first approximation in a spherically symmetric manner. The number of oxygen atoms present in layer of pure oxide 10^{-4} mm thick forming the surface of a sphere of radius 1 cm is about 7×10^{18}. The number of oxygen atoms needed to form the observed oxide layer is thus four to five orders of magnitude larger than the number present in the whole apparatus.

A supply of oxygen from the walls not out-

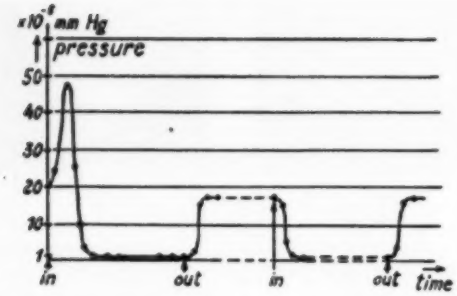


FIG. 6. Pressure course during evaporation, interrupted twice; Recorded by the AEG ionization gauge IM1. Between the individual evaporation at least an interval of one minute.

gassed and from the atmosphere through the numerous gaskets of the apparatus must therefore be supposed. In order to explain the oxidation of the evaporated metal quantitatively, the participation of both oxygen sources must probably be invoked. Since such large quantities of oxygen are bound by the evaporated metal, the pressure in the apparatus must decrease as long as oxide is being formed.

This effect was actually observed by an AEG ionization gauge type IM1 (see Fig. 6). The oxidation of metal vapors seems to be an effect accompanying the well-known getter action of evaporated metals.

The obtained results are the more significant since the mean free path of a Mg atom in air at 10^{-4} mm Hg as calculated from the kinetic theory diameter of the Mg, N_2, and O_2 molecules leads to much larger values than the distance from the atom source to the intercepting sheet.

At the beginning of the evaporation, the magnesium atoms enter a space containing O_2 molecules. The metal atoms, being of course in the nascent state, remove the oxygen molecules from the space around the atom source, binding them as metal oxide with a probability that will be in the first approximation inversely proportional to the square of the distance from the source. Given constant evaporation rate, the space becomes more and more depleted in oxygen molecules, and, proportionately with this depletion, the chance of a metal atom to get unchanged to the intercepting plane increases. The space surrounding the sources is—so to speak—gettering itself free of oxygen by means of the reacting metal vapor. Simultaneously with this effect, other O_2 molecules are diffusing toward the atom source from external regions, according to the diffusion coefficient and to the concentration gradient at each point in space. The mathematical solution of the whole calculation will be facilitated by the fact that the whole problem is spherically symmetric.

For every evaporation rate an equilibrium is attained between the disappearance of oxygen molecules and the simultaneous infusion.[7]

Several times heretofore the appearance of metal oxides has been observed while evaporating metals. J. A. Darbyshire[8] evaporated Kemet getters, consisting of 65 percent Mg, 20 percent Sr, and 15 percent Ba, in oxide cathode tubes. He found, also by electron diffraction, only magnesium oxide on the surface of the cathode.*** Since these investigations were carried out in a metal diffraction system sealed with rubber gaskets, the pressure was likely of the same range as in the present investigation. R. Beeching[9] reports that he had found the patterns of Al_2O_3, Cr_2O_3, and Cu_2O on layers of some hundredths of a micron thickness, after having evaporated the metals in a vacuum of 10^{-4} mm Hg. The metal lattice pattern does not appear until the layers become thicker. The same effect was observed by O. Beischer.† Beeching also found the pressure drop accompanying evaporation, but his explanation of the growing oxide layers seems to be insufficient.

In very thin evaporated layers, a smaller density is often found than that for the compact material; possibly this phenomenon is connected with the appearance of the oxide in the very thin films. In addition, W. Walkenhorst's[10] statement may be noted, namely, that slower evaporation speed is followed by lower densities of the layer, the difference amounting to 30 percent.

According to I. B. Nathanson and C. L.

---

[7] H. Stahl and S. Wagener, Zeits. f. tech. Physik **24**, 280 (1943).

[8] J. A. Darbyshire, Proc. Phys. Soc. **50**, 638 (1938).

*** Strontium oxide would not become visible because SrO also forms the cathode surface. (See H. Gaertner, Phil. Mag. **19**, 82 (1935); J. A. Darbyshire, Proc. Roy. Soc. **A50**, 635 (1935).) Barium evaporates noticeably at much higher temperatures than magnesium and strontium.

[9] R. Beeching, Phil. Mag. **22**, 942 (1936).

† Private communication.

[10] W. Walkenhorst, Dissertation (Polytechnic Academy, Hanover, 1940).

Bartberger,[11] the optical behavior of thin coats resulting when the metals were sputtered in hydrogen points to the existence of surface oxides.

The expected drop in pressure of the testing compartment caused by insufficiently rapid infusion was observed by R. Sewig[12] in the case of ionization gauges. The removal of the active molecules from the space between the cathode and the grid does not occur because of binding by metal vapor, however, but by the ionization and subsequent withdrawal of ions by the electric field. A similar case is known which occurs at higher pressures (10 to 12 mm Hg) in the manufacture of low pressure light tubes with a noble gas filling (argon-neon) and mercury droplets inside. It takes a very long time (10 min. to ½ hour) after the mercury drops are placed in the tube until the blue mercury light is emitted from the entire length of tube.

If the layers are deposited in steps, i.e., by an interrupted evaporation, the above-mentioned balance is disturbed and O_2 molecules refill the space around the atom source. At the beginning of each of the several partial evaporations from the same heating coil the initial oxide film must occur. Step-by-step layers are expected to contain, therefore, intermediate oxide layers.

These intermediate oxide layers were exhibited with magnesium. A one-micron thick layer was evaporated on outgassed bases, showing the space lattice pattern of the metal together with faint traces of 002 and 022 of MgO. Some time (10 min.) afterwards, another layer of 0.5 to 1×10^{-4} mm thickness was evaporated on it; the diagrams exhibited now only the pattern of MgO. The thickness of such intermediate oxide layers would depend on the initial vacuum.

Evidences of such inhomogeneities in stepwise-produced nickel coats were stated by R. Riedmiller[13] who examined the intensity distribution of x-ray diffractions. The diffracted intensities could be explained by assuming that both the total layer and the successive partial layers produce interference phenomena which superimpose on each other. W. Walkenhorst[14] encountered the same effect in measuring the reflective power and transmission of aluminum layers obtained by stepwise evaporations. Such layers showed an entirely anomalous behavior as contrasted to compact ones. The layers were supposed to be inhomogeneously built up and the secondary planes between the partial layers to be abrupt transition planes of the optical constants. About the nature of the separating layers Walkenhorst was not able to say anything definite.

In both instances the vacuum of about 10^{-4} mm was not interrupted between the partial evaporations. It may be noted, in addition, that J. Strong and E. Gaviola[15] made aluminum coats in a glass bell jar at 10^{-5} to 10^{-4} mm pressure also. The evaporation took place in a step-by-step manner; however, the vacuum was broken several times. The authors regarded the individual partial layers as being separated from each other by extremely thin alumina films.

V. SUMMARY

In a kinetic vacuum apparatus, certain metals (Be, Mg, Ca, Sr, Al, Mo, Ni) were evaporated and then condensed on outgassed sheet metal bases, or on glass or mica, and the structure of the resulting coats examined by electron diffraction. If the evaporated mass is so minute that films of thickness 0.5 to 1×10^{-4} mm are formed, the diffraction patterns prove the presence of metal oxides. In the uninterrupted evaporation of larger masses of Be, Al, Mo, and Ni, the diffraction pattern of the metal space lattice is evident, these layers being thicker than 10^{-4} mm. Under the same conditions, the getter metals, Mg, Sr, and Ca, never exhibit the diffraction pattern of the metal space lattice alone, but the metal pattern is always accompanied by the most intense rings of the oxide space lattice. No sign of the simultaneous occurrence of the corresponding nitrides was ever found.

During evaporation the vacuum improves, and hence a connection is suggested between the appearance of surface oxide and the known getter

[11] I. B. Nathanson and C. L. Bartberger, J. Opt. Soc. Am. **30**, 92 (1940); see also H. Zahn and J. Kramer, Zeits. f. Physik **86**, 417 (1933).

[12] R. Sewig, Zeits. f. tech. Physik **12**, 221 (1931). *Note added in proof:* J. Hillier (J. App. Phys. **19**, 229 (1948)) reports the same phenomenon in the case of electron microscope.

[13] R. Riedmiller, Ann. d. Physik (5) **20**, 397 (1934).

[14] W. Walkenhorst, Zeits. f. tech. Physik **22**, 19 (1941); Dissertation (Polytechnic Academy, Hanover, 1940), p. 22.

[15] J. Strong and E. Gaviola, J. Opt. Soc. Am. **26**, 157 (1936).

phenomenon of metal vapors. The metal first evaporated seems to serve always to getter the space around the metal vapor source, the bound oxygen appearing as metal oxide on the intercepting plane.

The results not only are confirmed by other investigations, but also account for some effects observed on stepwise evaporated films.

A further paper will concern phenomena associated with heated metal layers, the effects of thich also seem due to residual gases in the vacuum system.

The Thermal Behavior of Evaporated Layers in Vacuum Devices*

H. A. STAHL**
Laboratory for Material Physics, Telefunken Tube Factory, Berlin, Germany
(Received March 28, 1947)

Some layers presenting the space lattice pattern of the metal (Be, Mg, Al, Mo, or Ni) were heated for constant periods. Beginning with a distinctly marked threshold temperature, the interference rings of the oxide space lattice appeared while, simultaneously, the metal space lattice pattern diminished, entirely disappearing with another higher lying temperature. The oxidation temperature ranges (for ten-minute heating periods) of Be, Mg, and Al were measured. With glucinum it is equal to that within which, Kollath stated, there was a strong increase of the emissivity of beryllium secondary emission cells.

An elementary calculation gives an inferior time limit to oxidation of the observable surface layer. The surface layer of polished metals is considered; the high secondary emissivity in the case of glucinum, together with the usually supposed Beilby layer giving only two broad diffuse diffraction rings, points to the existence of a random liquid-like network of BeO molecules on the surface of polished beryllium secondary emission cells.

I. PROBLEM

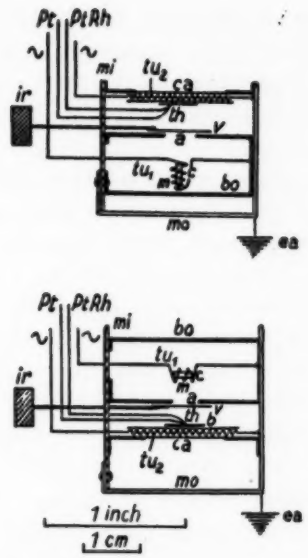
FOR several technical purposes, especially in the manufacture of secondary emission or photoelectric cells, sensitive surface layers are heated to rather high temperatures in order to submit them to a required heat treatment. Usually these layers have been produced by evaporation of a metal sample or by sputtering from a metal disk.

The present investigation, a continuation of the preceding one, was suggested by two observations, independent of each other. In making certain types of tubes, another getter besides the Fe-Ba getter is evaporated near the cathode. This additional getter consists essentially of magnesium. The very first portion of the metal vapor is oxidized under common production conditions, as shown in the preceding paper. But even if layers thicker than one micron are condensed on the cathode surface to give coats which could be expected to be metallic, electron diffraction patterns exhibit, in all cases, the space lattice diagram of magnesium oxide. The cathodes, removed and examined immediately afterwards, were, of course, exposed to the atmosphere for one or two minutes. On the other hand, magnesium layers, stored for several hours in air, showed the same pattern as those which were made in the diffraction apparatus itself and examined without interrupting the vacuum. It must be emphasized that the layers deposited on the oxide cathode are subjected to temperatures corresponding to dull red emission while the cathode is in operation. It became of interest, therefore, to inquire into the phenomena which occur when evaporated layers are heated to high temperatures.

The range of vacuum is the same as that in the preceding paper. In contrast, R. Kollath's investigations of the secondary emission of

* This document is based on work performed 1942 to 1943 within the research program of a foreign worker physicist in Germany. The printing was delayed by the post-war conditions in Europe.

** Now at the U.S.-Army Technical Information and Education School, Stuttgart-Vaihingen, Germany. Ex-member of the Czech Association of Physicists and Mathematicists, and of the Czechoslowak Testing Corporation, both Prague, and of the German Society for Technical Physics, Berlin.



FIGS. 1 and 2. Designs for thermal treatments of condensed metal coats mounted between a molybdenum bridge, *mo*, and a mica wing, *mi*. The metal sample, *m*, is evaporated from the tungsten coil, *c*, through the aperture, *a*, the valve, *v*, holding back the first evaporation rate. The base, *b*, is outgassable by an alumina-coated tungsten coil, tu_2. The thermocouple, *th*, is welded to the base *b*.

beryllium coats obtained by evaporation were carried out in highly outgassed, static vacuum systems. The cells were blown to high vacuum tubing, incorporating a charcoal trap, and could, therefore, be evacuated to 10^{-7} mm Hg, corresponding to a vacuum factor of 10^{-6}. Kollath[1] found that the secondary emission factor of such beryllium coats (on tungsten, molybdenum, tantalum, or beryllium as bases) is very much increased when the layer is subjected to a heat treatment called "formation." While heating to not more than 700°C is not accompanied by any remarkable phenomena, the secondary emission rises rapidly after heating the coat to more than 750°C, attaining the maximum value quoted for compact polished beryllium disks. Formation takes place in the temperature interval from 750 to 900°C since further temperature increase is not followed by any additional secondary emission. As an explanation of this effect Kollath can only propose structure changes of unknown type: "The suddenness of these changes after passing a definite threshold temperature leads to the conclusion of a displacement of the beryllium atoms in the layer, evidently influencing the secondary emission."

In his later papers he differs from these hypothetical displacements and confesses the possibility of oxidation.[2]

II. EXPERIMENTAL ARRANGEMENT

Thus it was of interest to investigate the surface of metallic layers obtained by evaporation, before and after heat treatment in vacuum, by means of electron diffraction. The same apparatus and high voltage plant were used as in the experiments described previously. The acceleration voltage was again 80 kv, corresponding to a de Broglie wave-length of 0.0417A.

The bases, *b*, on which the films were deposited were outgassed by the heating coil, tu_2 (see Fig. 1). Nickel-carbonyl, platinum, and occasionally iron or molybdenum were used as bases. The results did not differ perceptibly. An alumina-coated tungsten coil, encased in a nickel casing, *ca*, was again used as a heater, as in indirectly heated vacuum tubes. The layers were first made in the arrangements shown in Figs. 1 and 2, and then submitted to the thermal treatment immediately afterwards. Diffraction patterns were than taken at grazing incidence, maintaining the vacuum. The evaporated masses were always large enough to cause the mica window, *w*, to become opaque, indicating a layer thickness of one micron or more. Close to the spot examined, a le Chatelier thermocouple was welded, thus providing an accurate method of measuring the temperature during heat treatment. If may be noticed that the electrons scattered by the last diaphragm yielded an enlarged silhouette of the couple, permitting the subsequent calculation of the distance from the couple to the particular spot examined. The platinum and platinum-rhodium wires were conducted through a seal-out of the diffraction apparatus before being connected to copper, in order to avoid inaccuracies caused by undesired thermocurrents.

To allow evaporation under better, and mainly under static, vacuum conditions, tests were also

[1] R. Kollath, *Jahrbuch des Forschungsinstitutes der Allgem. Elektricitäts-Gesellschaft* (1936/1937), Vol. V, p. 61; Naturwiss. **26**, 60 (1938); Ann. d. Physik (5) **33**, 285 (1938); **39**, 76 (1940); Zeits. f. tech. Physik **21**, 331 (1940); Physik. Zeits. **41**, 576 (1940); Fernsehen und Tonfilm, No. 4/5, 14 (1941).

[2] See also G. Gille, Zeits. f. tech. Physik **22**, 228 (1941); I. B_3N_2 is marked (poorly visible in the reproduction).

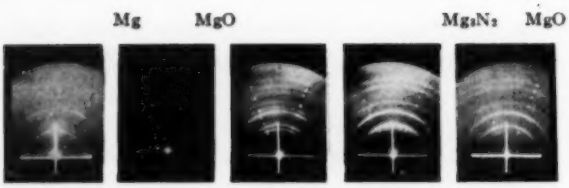


FIG. 3. Transformation series Mg→MgO (+Mg_3N_2). At the left of the pictures Mg rings dotted, at right MgO rings, respectively. In the last picture the strongest ring of Mg_3N_2 pointed, but in the reproduction not visible.

conducted with layers made outside the diffraction apparatus. The evaporation furnace was placed under a glass jar or a standard glass cone of about 7 cm in diameter, the vacuum being 10^{-5} mm Hg. The coated sheet metal pieces were exposed to the atmosphere for less than two minutes. Separately undertaken control experiments proved that this short contact with the atmosphere had no effect on the results obtained. Except for the known getter metals, the surfaces exhibited the pattern of the metal space lattice before heat treatment.

III. RESULTS

Magnesium, beryllium, and aluminum were first investigated; rather qualitative experiments were performed with molybdenum and nickel, in addition. The base materials were the same as those used in the preceding paper.

The results are uniform: every metal lattice undergoes a radical change above a threshold temperature specific for each metal; at the same time residual gases in the apparatus seem to become chemically bound. The usual heat treatments of the metal films lasted ten minutes, but layers were also heated repeatedly and gradually to higher and higher temperatures until the surface structure underwent no further change. The results obtained are as follows.

A. Magnesium

A film exhibiting the diffraction pattern of the metal space lattice (with 002 and 022 of MgO also visible in traces) was heated to 300°C and then to 700°. In the first case the diagram remained unmodified compared with the untreated layer; in the latter case the metal lattice disappeared entirely and was replaced by a face-centered cubic lattice of constant 4.20A. In addition, a weak additional ring appeared. The complete pattern was ascribed to MgO; the additional ring proved to be the most intense interference of Mg_3N_2.[3]

The temperature range within which the magnesium pattern disappears and that of magnesium oxide emerges now had to be evaluated. After heating the coat to 320°C, the 002 and 022 rings of MgO were considerably stronger than they were originally. The oxidation appears to be incomplete even after heating to 370 and 430°C, the most intense ring of the metal lattice diagram remaining visible. Only after the temperature was increased to 500°C was the oxidation complete, the complete ring system of MgO and the single Mg_3N_2 ring appearing (see Fig. 3).

Magnesium was also condensed on glass slips. The temperatures found were considered as only approximate since the thermocouple was pressed between the glass and the heating case effecting rather poor thermal contact of the slip. The results, nevertheless, seem entirely in accord with those obtained with metal bases for deposition.

Magnesium films on outgassed nickel or iron bases showed no increase in relative intensity of the MgO interference lines after exposure to the atmosphere for one hour.

B. Beryllium

Layers exhibited behavior entirely analogous to that of magnesium. Metal films heated to 300°C showed no changes of surface structure, but after being heated to 800°C only the BeO space lattice in addition to another ring, was exhibited. According to M. v. Stackelberg and R. Paulus,[4] this ring is the second most intense (044) in the pattern belonging to Be_3N_2. The most intense interference (222) coincides with the 010 of BeO and its presence could have been confirmed only by photometering the patterns. This was not undertaken, all the more since Stackelberg and Paulus' statements represent x-ray intensities.

[3] J. D. Hanawalt, H. W. Rinn, and L. K. Frevel, Ind. Eng. Chem. **10**, 457 (1938).
[4] M. v. Stackelberg and R. Paulus, Zeits. f. physik. Chemie **B22**, 315 (1933).

The more precise investigation of the oxidation process indicated that no oxide is formed even after heating to 700°, the oxide pattern emerging only after a temperature of 740° is attained (the five most intense rings of the metal diagram are still visible at these temperatures). After heating to 800° the metal space lattice predominates, only the most intense interference of the metal lattice persisting. The latter disappears entirely upon heating to 900°C (see Fig. 4).

Evaporated beryllium layers which exhibited only the metal space lattice pattern immediately after deposition showed the same diagram even after having been stored 12 days in open air.

C. Aluminum

Aluminum behaved like magnesium and beryllium. It must, however, be remembered that aluminum oxide exists also in an amorphous phase. The first preliminary patterns showed that after heating to 420°C no noticeable oxidation had taken place. After heating to 500° the most intense interference (044) of γ-Al_2O_3 becomes perceptible. After being heated to 580° and 600° the three most intense rings of γ-Al_2O_3 (133, 004, 044) and 002, 022, and 224 of the metal space lattice may be seen. Evaporated films heated to 620° no longer exhibited any rings of the metal space lattice, the oxidation appearing to be complete.

Aluminum patterns exhibited no change, even after the sample having been stored in air for two weeks and more.

D. Molybdenum

Films were evaporated in the diffraction apparatus itself and investigated immediately afterwards. No thermocouple was attached to the intercepting sheet metal. Although before treatment the layer exhibited the space lattice diagram of metallic molybdenum, after being heated to a red glow the diagram corresponding to Mo_2O_3 (according to Hanawalt) appeared (see Fig. 5). The existence of this compound is doubtful, according to other authors.[5]

[5] J. D. Hanawalt, R. W. Rinn, and L. K. Frevel, Ind. Eng. Chem. **10**, 1496 (1938). See, however, H. Remy, *Lehrbuch der anorg. Chemie* (Akademische Verlagsgesellschaft Becker & Erler, Leipzig, 1942), Vol. II, p. 166; Gmelin's *Handbuch der anorg. Chemie* (Verlag Chemie Berlin, 1935), System-No. 53, p. 90.

E. Nickel

After having been heated to a bright red glow, nickel showed the diagram of a face-centered cubic lattice of lattice constant 4.17A, which belongs to NiO, according to Hanawalt, Rinn, and Frevel.

F. Summary of Results

In summarizing it can be said that all the metals investigated oxidize superficially within characteristic temperature intervals, the time of heat treatment having been, in all cases, about ten minutes. The beginning of these intervals was fixed by the appearance of the most intense interference of the metal oxide space lattice; the end by the disappearance of the most intense ring of the metal space lattice diagram. Table I gives these values for the metals investigated in greater detail. For molybdenum and nickel these temperatures lie between room temperature and a bright red glow.

IV. DISCUSSION

The results obtained clearly state that not only reactive alkaline earth metals like magnesium, but also other basic metals like aluminum, molybdenum, and nickel are not protected against surface oxidation upon being heated in a kinetic vacuum system of 10^{-5} to 10^{-4} mm Hg.

These results have been confirmed by other authors. R. W. Schmidt[6] investigated tungsten and molybdenum powders in the kinetic vacuum of an electron microscope. He found the presence of metal oxides clearly indicated. In the case of tungsten, some comparisons resulted in the confirmation of a highly oxidized compound, W_4O_{11}. Schmidt supposed the powder tempera-

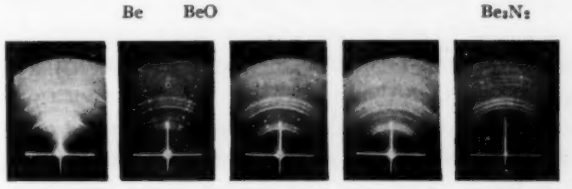


FIG. 4. The same for the transition Be→BeO (+Be_3N_2). In the last photogram the second most intense ring of Be_3N_2 is marked (poorly visible in the reproduction).

[6] R. W. Schmidt, Kolloid Zeits. **102**, 15 (1943).

Mo (a) Mo_2O_3 (b)

FIG. 5. Evaporated molybdenum layer, before (a) and after (b) glowing in kinetic vacuum systems.

ture to be increased greatly because of the energy imparted to the metal crystals by the electrons (1000°C or more). Because of the presence of traces of oxygen in the microscope, this high temperature manifests itself in the appearance of the metal oxide. H. Mahl[7] likewise finds that the swelling of cellulose fibers, often observed in electron microscopes, can probably be explained as direct combustion of the fiber.

The results on aluminum are verified by G. Hass[8] on the basis of optical measurements, by electron diffraction, and microscopy. He found that aluminum overcoats itself with a distinctly demonstrable amorphous oxide skin about 20A thick (increasing to 90A with age). More extensive oxidation occurs when the object is heated in open air to 400° C, the oxide skin cracking and crystallizing as γ-Al_2O_3. At 450° this process is well defined. The obvious coincidence of the temperature ranges stated by Hass with the present ones suggests the conclusion that the same phenomenon is occurring.

The behavior of beryllium is very remarkable. Kollath[1] finds certain peculiar phenomena in beryllium secondary emission cells in a static vacuum system of 10^{-7} mm Hg beginning at a distinctly defined temperature and obviously ending at another higher temperature. The striking coincidence of his temperature range with that of the present paper again confirms the assumption that Kollath's experiments are due to real surface oxidation, i.e., that the enormous growth of the layer's secondary emission factor is accompanied by a simultaneous formation of beryllium oxide on the surface, and this substance accounts for the large secondary emission.

[7] H. Mahl, Kolloid Zeits. **96**, 7 (1941); see also L. Wallner, *Melliand Textilberichte* (Heidelberg), No. 4/6.

[8] G. Hass, Ann. d. Physik (5) **31**, 245 (1938); Verh. d. D. Phys. Ges. **22**, 1 (1941); Optik **1**, 134 (1946); G. Hass and H. Kehler, Kolloid Zeits. **97**, 31 (1941).

Here it is also shown that seemingly uncontestable vacuum conditions do not preclude the contamination of reactive metal surfaces. The range of pressures in which metal surfaces are oxidizable seems to be very wide, at least from 760 mm (for instance, aluminum) down to 10^{-7} mm (for instance, beryllium).

In fact, H. Bruining[9] points out that heating of reactive metals in vacuum is a "dangerous beginning," and the beryllium surface must be supposed to react with the gases set free from the base plate. The surface layer of beryllium oxide is made probably by earlier experiments of H. Bruining and J. H. de Boer,[10] who succeeded, in the case of Al and Be, under extremely clean conditions, in removing the oxidized layer, thus regaining the original low secondary emission of the evaporated metal surface.

The oxidation rate of the surface layer which contributes to the diffracted intensity for 80-kv electrons can be estimated from elementary assumptions. Assuming that every impinging oxygen molecule will be bound by the metallic surface, provided the temperature is in the oxidation interval (Table I) or above it, the necessary time can be calculated for every given partial pressure of oxygen. The thickness of the layer is supposed to be 100A or 10^{-6} cm, the volume per cm^2 then 10^{-6} cm^3. The elementary cell of the magnesium space lattice (volume $46A^3$) contains two magnesium atoms; the quantity of magnesium atoms inside the surface layer is therefore

$$n_1 = 2(10^{-6}/46 \cdot 10^{-24}) = 4.35 \times 10^{16}\ \text{cm}^{-2}.$$

On the other hand, $z = (n_2/6)v$ molecules are impinging on each square centimeter per second, where n_2 is the number of molecules per cm^3 and v is the mean velocity. A gauged pressure of 1.5×10^{-4} mm corresponds to an oxygen partial pressure of 3×10^{-5} mm Hg, hence the equation

$$n_2 = (2.7 \cdot 10^{19}/760)3 \cdot 10^{-5} = 10^{12}$$

holds. The velocity v can be calculated from

$$\tfrac{3}{2} \cdot kT = \tfrac{1}{2} \cdot u \cdot v_{\text{eff}}^2$$

for $T = 293°\text{K}$, u representing the oxygen mole-

[9] H. Bruining, Physica **8**, 1161 (1941); *Die Sekundär-Elektronenemission fester Körper* (Verlagsbuchhandlung Julius Springer, Berlin, 1942), pp. 31–33.

[10] H. Bruining and J. H. de Boer, Physica **4**, 473 (1937).

cule mass, and

$$v = 0.922 v_{\text{eff}}.$$

We obtain therefore

$$v = 0.922(3kT/u)^{\frac{1}{2}}$$
$$= 0.922(3 \cdot 293 \cdot 1.38 \cdot 10^{-16}/32 \cdot 1.67 \cdot 10^{-24})^{\frac{1}{2}}$$
$$= 4.38 \cdot 10^{4} \text{ cm sec.}^{-1}.$$

The number of impinging O_2 molecules/cm² sec. is then

$$z = (10^{12}/6) \cdot 4.38 \cdot 10^{4} = 0.73 \cdot 10^{16},$$

or the quantity of bound atoms in the 100A layer

$$2z = 1.46 \cdot 10^{16}.$$

Under the mentioned assumptions, the layer is wholly oxidized after

$$t = n_1/2z = 4.35 \cdot 10^{16}/1.46 \cdot 10^{16} = 3 \text{ sec.}$$

The surface layer investigated is, however, between 100 and 1000A thick; on the other hand, a portion of the impinging O_2 molecules can be reflected. The calculated time represents, therefore, only the inferior time limit for oxidation.

The same calculation, carried out for beryllium for a total gauge pressure of 10^{-7} mm Hg, i.e., for an oxygen partial pressure of 2×10^{-8} mm, gives the figures

$$n_1 = 6.2 \times 10^{16}$$

and

$$2z = 10^{13}.$$

Hence, t will be 6.2×10^{3} seconds or 100 min.

Since Kollath finds, on the contrary, that the formation is already finished after ten minutes' heat treatment, some other factor must be operative. The real partial pressure over the surface at the time of formation may be higher than supposed by a factor of ten. In part, this occurs because the exhausting mechanism of the pump shifts the equilibrium O_2/N_2 in favor of the heavier oxygen.

That the gauge pressure of 10^{-7} mm really represents the true pressure in the bulb at the moment of formation may be considered improbable. To get reliable conditions in vacuum devices, all parts have to be outgassed by heating them at considerably higher temperatures than those later to be attained. With beryllium, this condition may not be met, since the metal sample must always be kept at a temperature below that at which evaporation takes place. Hence, the metal is not outgassed until the evaporation process itself is taking place. The gases set free would then be absorbed, at least in part, in the layer just deposited. The secondary emission layer can, therefore, be said to contain minute oxygen traces, even if the base plate had previously been completely outgassed. It is these traces which finally appear in the surface.

TABLE I.

Metal	Appearance of the strongest oxide ring °C	Disappearance of the strongest metal ring °C
Beryllium	700	900
Magnesium	320	500
Aluminum	450	620

V. REMARKS ON SECONDARY EMISSION MEASUREMENTS

If the above conclusions are correct, changes in the secondary emission efficiency of evaporated layers can be expected for other metals upon heating them above definite threshold temperatures. According to the results of the preceding paper, the evaporated layers must be presumed to be compact and homogeneous, i.e., made in a single evaporation process. In good agreement with this, E. G. Schneider[11] finds that upon heating such layers, higher secondary emission factors are obtained, and upon further evaporation of metal the lower ones occur again.

If the layer is extremely thin (100 to 1000A), a smaller increase in the emission factor may be expected after heat treatment, because of the existence of the excessively thin surface oxide, as shown in the preceding paper. It is worthy of mention that this expectation is well confirmed by Kollath's measurements on layers of various thicknesses. From his statements*** are calculated the ratios compiled in Table II. The quotient is shown to be monotone increasing with increasing layer thickness, which effect may be ascribed to the increasing portion of oxidizable metal in the effective surface layer.

Kollath mentions that the very high secondary emission coefficients obtained by the formation process may also be reached on polished beryl-

[11] E. G. Schneider, Phys. Rev. **54**, 185 (1938).
*** Figure 7 in Ann. d. Physik (5) **33**, 292 (1938).

TABLE II.

Layer	Secondary emission factor after formation / Secondary emission factor before formation
Very thin	43/8.5 = 5.1
Medium	55/6 = 9.2
Thick	63/5.5 = 11.5

lium disks. This effect may be expected, according to the results of the present paper. In the case of easily oxidizable metals like beryllium, magnesium, or copper, it has been stated by electron diffraction researchers that the surface becomes oxidized in the polishing process. According to F. B. Bowden and T. P. Hughes,[12] surface temperatures as high as 600 to 800°C occur. These temperatures, indeed, in the case of magnesium and aluminum, are higher than the threshold temperatures quoted in Table I; for beryllium the values are nearly equal. Thus, the surface of polished mirrors probably consists of beryllium oxide, but it is uncertain whether the beryllium oxide space lattice is present.

On polished surfaces an amorphous liquid-like "Beilby layer" is generally supposed to exist, having nothing in common with the space lattice of the normal crystalline substance, and consisting of a simple network of statistically distributed atoms. On the other hand, the secondary emission must be attributed not only to the atoms themselves, but also to the lattice—as resulted from the investigations of S. R. Rao[13] on the one hand, and G. Bekow[14] on the other.

A beryllium oxide network on the surface of a polished beryllium disk cannot be expected to yield the same secondary emission as a normal beryllium oxide space lattice. Considering the Beilby effect, some other typical surface properties of polished metals must also not be presumed similar to evaporated layers, even if the polished surface is regarded as metallic. G. Hass[15] accordingly found that evaporated silver layers differ widely optically from polished mirrors.

The secondary emission of a polished beryllium disk, as investigated by P. L. Copeland[16] and R. Kollath and his co-workers,[1,2] would be attributed, therefore, not to the crystalline beryllium lattice, but either to a more or less disturbed BeO space lattice or to a BeO network.

[12] F. B. Bowden and T. P. Hughes, Proc. Roy. Soc. **A160**, 575 (1937).
[13] S. R. Rao, Proc. Roy. Soc. **A128**, 57 (1930).
[14] G. Bekow, Physik. Zeits. **42**, 144 (1941).

VI. SUMMARY

Layers of beryllium, magnesium, aluminum, molybdenum, and nickel 10^{-3} mm thick or more were evaporated upon outgassed bases in a kinetic vacuum system of 10^{-5} to 10^{-4} mm Hg, and the thermal behavior of these coats was investigated by electron diffraction. For the three first-quoted metals, characteristic threshold temperatures were found above which the space lattice of the metal oxide appears after a heat treatment of ten minutes. For a uniform heating time the interference rings of the original metallic space lattice disappear at another different temperature, likewise characteristic for each metal.

Both these end values for the oxidation interval are reported for Be, Mg, and Al; for the other metals they lie between room temperature and red heat.

Layers remaining exposed to the air for long times (days or even weeks) below the inferior threshold temperature, exhibit no perceptible changes in their interference patterns.

From the satisfactory coincidence of the temperature intervals obtained by this method with that one following from secondary emission measurements, additional conclusions are drawn upon the surface structure of high vacuum secondary emission cells on the one hand, and upon the residual gases in static vacuum systems on the other.

[15] G. Hass, Ann. d. Physik (5) **31**, 261 (1938).
[16] P. L. Copeland, Phys. Rev. **92**, 46 (1934).

Stress-Strain Rate Relations for Anisotropic Plastic Flow*

JOHN E. DORN
Division of Mineral Technology, University of California, Berkeley, California
(Received April 29, 1948)

A simple theory for stress-strain rate analyses during plastic flow of anisotropic sheet metals is proposed for the work hardening range. The effect of orientation on the yield strength in simple tension and biaxial tension was discussed for various types of symmetry. The theory appears to be approximately correct for mild steel plates that exhibit planar isotropy.

INTRODUCTION

PREFERRED grain orientation and banding of microconstituents frequently cause wrought metals to have anisotropic plastic properties. Although extensive investigations have been made on the laws of isotropic plastic flow, no method of analysis has yet been published for correlating the stresses and strain rates in anisotropic materials. Such an analysis is presented here for problems of planar stressing pertinent to sheet metals.[1-3] The procedures, however, are sufficiently general to permit extension of the method to the three dimensional case.

STRESS-STRAIN RATE RELATIONS

The two major assumptions of the generalized stress-plastic strain rate relations are extensions of experimental observations made on the more thoroughly investigated examples of isotropic plastic flow.

1. The plastic strain rates are functions of the stresses.[4-6]
2. To a first approximation the components of the strain rates are linear functions of the stresses.[7-10]

Therefore the six plastic strain rate-stress equations become[6,9]

$$\begin{aligned}\dot{\epsilon}_{xx}&=\dot{A}_{11}\tau_{xx}+\dot{A}_{12}\tau_{yy}+\dot{A}_{13}\tau_{zz}\\&\quad+\dot{A}_{14}\tau_{xy}+\dot{A}_{15}\tau_{yz}+\dot{A}_{16}\tau_{zx},\\\dot{\epsilon}_{xy}&=\dot{A}_{41}\tau_{xx}+\dot{A}_{42}\tau_{yy}+\dot{A}_{43}\tau_{zz}\\&\quad+\dot{A}_{44}\tau_{xy}+\dot{A}_{45}\tau_{yz}+\dot{A}_{46}\tau_{zx},\end{aligned}\qquad(1)$$

where $\dot{\epsilon}_{xx}$ = the strain rate on the fiber instantaneously in the x coordinate direction, $\dot{\epsilon}_{xy}$ = the shear strain rate between fibers instantaneously in the x and y directions, τ_{xx}, etc. = the instantaneous components of the normal stresses, τ_{yy}, etc. = the instantaneous components of the shear stresses, and $\dot{A}_{kl}(k, l=1 \text{ to } 6)$ = coefficients of plasticity having dimensions of a strain rate over a stress.

Although

$$\dot{\epsilon}_{ij}=(d\epsilon_{ij}/dt)(i, j=x, y, z)$$

was called a strain rate, actually two meanings will be ascribed to the variable t.[9] When the plastic properties are sensitive to strain rate effects t will be simply the time. But when the plastic properties are practically independent of strain rate and are primarily functions of the preceding deformation, as in the work hardening range for metals, t will be a parameter that determines the extent of the deformation. Therefore, in a torsion test of a metal bar in the work

* This paper is based on some of the data reported to the ONR in the first three references. The author acknowledges the ONR sponsorship of these investigations and expresses his appreciation for authorization to publish these data.

[1] J. E. Dorn, *Stress-strain relations for the plastic deformation of metals*, Third Progress Report to Office of Naval Research (June, 1947).

[2] T. E. Tietz, C. H. Avery, and J. E. Dorn, *Investigations on the validity of an idealized theory of plastic flow for rolled mild steels*, Fourth Progress Report to the Office of Naval Research (November, 1947).

[3] J. E. Dorn, *Work hardening of anisotropic sheet metals*, Fifth Technical Report to the Office of Naval Research (February 16, 1948).

[4] W. Lode, "Der Einfluss der Mittleren Hauptspannung auf das Fliessen Metalle," Forsch. a.d. Gebiete d. Ingenierwissens No. 303 V.D.I. (Verlag Julius Springer Berlin, 1928).

[5] G. I. Taylor and H. Quinney, "The plastic distortion of metals," Phil. Trans. Roy. Soc. London **A230**, 323-362 (1931).

[6] Jelinek, Latter, Thomsen, and Dorn, "Plastic Flow in Metals," O.P.R.D. Report No. 3218 (May 1945).

[7] A. Nadai, *Plasticity* (McGraw-Hill Book Company, Inc., New York, 1931).

[8] W. Prager, "Strain hardening under combined stresses," J. App. Phys. **6**, 837-840 (1945).

[9] J. E. Dorn and E. G. Thomsen, "The ductility of metals under general conditions of stress and strain," Trans. Am. Soc. Met. (1947).

[10] D. M. Cunningham, E. G. Thomsen, and J. E. Dorn, "Plastic flow of a magnesium alloy under biaxial stresses," Am. Soc. Test. Mat. **47**, 546-553 (1947).

hardening range, t is the twist of the bar in radians per degree. Such a dual interpretation is useful because it permits analyses of both flow rate and deformation phenomenon by the same analytical procedures.

The coefficients of plasticity, unlike the coefficient of viscosity or the moduli of elasticity, are not constants; they depend upon the previous history of flow and deformation; therefore, they are not symmetrical (in general $\dot{A}_{kl} \neq \dot{A}_{lk}$) and the 36 coefficients of plasticity may differ from each other.

For planar plastic flow, Eqs. (1) reduce to

$$\begin{aligned} \dot{\epsilon}_{xx} &= \dot{A}_{11}\tau_{xx} + \dot{A}_{12}\tau_{yy} + \dot{A}_{14}\tau_{xy}, \\ \dot{\epsilon}_{yy} &= \dot{A}_{21}\tau_{xx} + \dot{A}_{22}\tau_{yy} + \dot{A}_{24}\tau_{xy}, \\ \dot{\epsilon}_{xy} &= \dot{A}_{41}\tau_{xx} + \dot{A}_{42}\tau_{yy} + \dot{A}_{44}\tau_{xy}. \end{aligned} \tag{2}$$

X-ray diffraction studies of grain orientation and microscopic investigations of banding reveal that all sheet metals exhibit a plane of symmetry.[11] If the x axis is taken in the cross-rolling direction, the y axis in the rolling direction, the $y-z$ plane is a plane of symmetry. Since Eqs. (2) must preserve their original equality after reflection across the plane of symmetry, as obtained by the transformation

$$x' = -x, \quad y' = y, \quad z' = z,$$

the coefficients $\dot{A}_{14}$, $\dot{A}_{24}$, $\dot{A}_{41}$, $\dot{A}_{42}$ must vanish. Then the equations,

$$\begin{aligned} \dot{\epsilon}_{xx} &= \dot{A}_{11}\tau_{xx} + \dot{A}_{12}\tau_{yy}, \\ \dot{\epsilon}_{yy} &= \dot{A}_{21}\tau_{xx} + \dot{A}_{22}\tau_{yy}, \\ \dot{\epsilon}_{xy} &= \dot{A}_{44}\tau_{xy}, \end{aligned} \tag{3}$$

represent the stress-plastic strain rate relations for the lowest symmetry commonly found in sheet metals. Sheet metals also have an $x-z$ plane of symmetry but no additional simplification of the equations of plasticity results from this fact. Other symmetries, however, further reduce the number of independent coefficients of plasticity.

DIRECTIONAL PROPERTIES

The variation of plastic properties with orientation in the sheet is due to changes of the coefficients of plasticity with orientation and can be determined by methods analogous to those used in the anisotropic theory of elasticity.[12,13] To evaluate such changes, consider a rotation of axis θ degrees about the z axis. A point at x, y of the original system will then have position,

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta, \end{aligned} \tag{4}$$

in the new coordinate system. After transformation, the strain rates are

$$\begin{aligned} \dot{\epsilon}_{x'x'} &= \dot{\epsilon}_{xx}\cos^2\theta + \dot{\epsilon}_{yy}\sin^2\theta + \dot{\epsilon}_{xy}\sin\theta\cos\theta, \\ \dot{\epsilon}_{y'y'} &= \dot{\epsilon}_{xx}\sin^2\theta + \dot{\epsilon}_{yy}\cos^2\theta - \dot{\epsilon}_{xy}\sin\theta\cos\theta, \\ \dot{\epsilon}_{x'y'} &= 2(\dot{\epsilon}_{yy} - \dot{\epsilon}_{xx})\sin\theta\cos\theta + \dot{\epsilon}_{xy}(\cos^2\theta - \sin^2\theta), \end{aligned} \tag{5}$$

and the stresses are related by

$$\begin{aligned} \tau_{xx} &= \tau_{x'x'}\cos^2\theta + \tau_{y'y'}\sin^2\theta - \tau_{x'y'}\sin\theta\cos\theta, \\ \tau_{yy} &= \tau_{x'x'}\sin^2\theta + \tau_{y'y'}\cos^2\theta + 2\tau_{xy}\sin\theta\cos\theta, \\ \tau_{xy} &= (\tau_{x'x'} - \tau_{y'y'})\sin\theta\cos\theta \\ &\qquad + \tau_{x'y'}(\cos^2\theta - \sin^2\theta). \end{aligned} \tag{6}$$

Accordingly, the stress-plastic strain rate equations become

$$\begin{aligned} \dot{\epsilon}_{x'x'} &= \dot{A}_{11}'\tau_{x'x'} + \dot{A}_{12}'\tau_{y'y'}, \\ \dot{\epsilon}_{y'y'} &= \dot{A}_{21}'\tau_{x'x'} + \dot{A}_{22}'\tau_{y'y'}, \\ \dot{\epsilon}_{x'y'} &= \dot{A}_{44}'\tau_{x'y'}, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \dot{A}_{11}' &= (\dot{A}_{11}\cos^2\theta + \dot{A}_{21}\sin^2\theta)\cos^2\theta \\ &\quad + (\dot{A}_{12}\cos^2\theta + \dot{A}_{22}\sin^2\theta)\sin^2\theta \\ &\quad + \dot{A}_{44}\sin^2\theta\cos^2\theta, \\ \dot{A}_{12}' &= (\dot{A}_{11}\cos^2\theta + \dot{A}_{21}\sin^2\theta)\sin^2\theta \\ &\quad + (\dot{A}_{12}\cos^2\theta + \dot{A}_{22}\sin^2\theta)\cos^2\theta \\ &\quad - \dot{A}_{44}\sin^2\theta\cos^2\theta, \\ \dot{A}_{21}' &= (\dot{A}_{11}\sin^2\theta + \dot{A}_{21}\cos^2\theta)\cos^2\theta \\ &\quad + (\dot{A}_{12}\sin^2\theta + \dot{A}_{22}\cos^2\theta)\sin^2\theta \\ &\quad - \dot{A}_{44}\sin^2\theta\cos^2\theta, \\ \dot{A}_{22}' &= (\dot{A}_{11}\sin^2\theta + \dot{A}_{21}\cos^2\theta)\sin^2\theta \\ &\quad + (\dot{A}_{12}\sin^2\theta + \dot{A}_{22}\cos^2\theta)\cos^2\theta \\ &\quad + \dot{A}_{44}\sin^2\theta\cos^2\theta, \\ \dot{A}_{44}' &= 4(\dot{A}_{11} - \dot{A}_{21} - \dot{A}_{12} + \dot{A}_{22})\sin^2\theta\cos^2\theta \\ &\quad + \dot{A}_{44}(\cos^2\theta - \sin^2\theta)^2. \end{aligned} \tag{8}$$

HIGHER SYMMETRIES

If sheet metal exhibits pronounced banding of the microconstituents, the highest possible symmetry is that of $y-z$ and $x-z$ planes of symmetry. But in the absence of banding, the symmetry is that determined by the pole figure of grain

[11] G. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., 1943).

[12] W. Voigt, "Lehrbuch der Kristallphysik," Math. Wissensch. Leipsig (1910).

[13] A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity* (Dover Publications, New York, 1944).

orientations. Occasionally the pole figures for sheet metals exhibit the symmetry resulting from superposition of a fourfold axis of symmetry on $y-z$ and $x-z$ planes of symmetry. Equations (8) a show that the form of Eqs. (3) are preserved after rotation of 90° if

$$\dot{A}_{11}=\dot{A}_{22} \quad \text{and} \quad \dot{A}_{12}=\dot{A}_{21}. \tag{9}$$

Consequently, the equations of plasticity for sheet metal having a fourfold axis of symmetry are

$$\begin{aligned} \dot{\epsilon}_{xx}&=\dot{A}_{11}\tau_{xx}+\dot{A}_{12}\tau_{yy},\\ \dot{\epsilon}_{yy}&=\dot{A}_{12}\tau_{xx}+\dot{A}_{11}\tau_{xx},\\ \dot{\epsilon}_{xy}&=\dot{A}_{44}\tau_{xy}. \end{aligned} \tag{10}$$

It is now known that mild steel plate, although not completely isotropic, may exhibit identical plastic properties for all directions in the plane of the sheet.[14] Reference to Eqs. (8) reveals that such planar isotropy demands, in addition to conditions of Eqs. (9), that

$$2\dot{A}_{11}-2\dot{A}_{12}-\dot{A}_{44}=0. \tag{11}$$

For planar isotropy, therefore,

$$\begin{aligned} \dot{\epsilon}_{xx}&=\dot{A}_{11}\tau_{xx}+\dot{A}_{12}\tau_{yy},\\ \dot{\epsilon}_{yy}&=\dot{A}_{12}\tau_{xx}+\dot{A}_{11}\tau_{yy},\\ \dot{\epsilon}_{xy}&=2(\dot{A}_{11}-\dot{A}_{12})\tau_{xy}. \end{aligned} \tag{12}$$

If the sheet metal is completely isotropic, stressing in the cross-rolling direction should induce equal strain rates in the thickness and rolling directions. Therefore, for this specialization

$$\begin{aligned} \dot{\epsilon}_{xx}&=\dot{A}_{11}\tau_{xx},\\ \dot{\epsilon}_{yy}&=\dot{A}_{12}\tau_{xx},\\ \dot{\epsilon}_{zz}&=\dot{A}_{12}\tau_{xx}. \end{aligned} \tag{13}$$

Inasmuch as the density of metals does not change appreciably with plastic deformation,[15] the equation of continuity,

$$\dot{\epsilon}_{xx}+\dot{\epsilon}_{yy}+\dot{\epsilon}_{zz}=0, \tag{14}$$

may be applied and, therefore

$$\dot{A}_{12}=-\dot{A}_{11}/2. \tag{15}$$

Consequently, for complete three-dimensional isotropy

$$\begin{aligned} \dot{\epsilon}_{xx}&=\dot{A}_{11}\{\tau_{xx}-\tfrac{1}{2}\tau_{yy}\},\\ \dot{\epsilon}_{yy}&=\dot{A}_{11}\{\tau_{yy}-\tfrac{1}{2}\tau_{zz}\},\\ \dot{\epsilon}_{xy}&=3\dot{A}_{11}\tau_{xy}. \end{aligned} \tag{16}$$

Equations (16) are those now commonly used to describe the plastic behavior of metals.[2,6,9,10] But these equations are seldom applicable to sheet metals because of their anisotropic characteristic. Obviously the appropriate selection of the equations of plasticity to be applied to sheet metals can be determined from pole figures and microscopic examination.

CONSTANT ANISOTROPY

When metals are subjected to large finite deformations their anisotropic characteristics

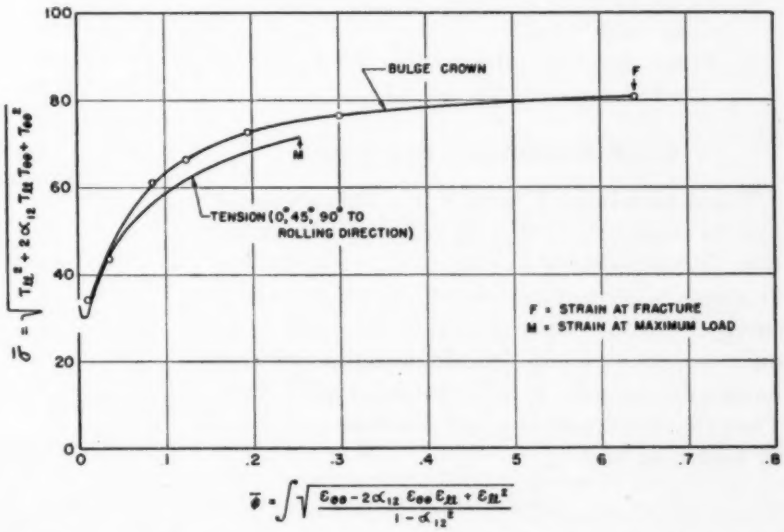


FIG. 1. Comparison of $\bar{\sigma}-\bar{\varphi}$ curves for bulging and tension of metal having planar isotropy when $\bar{\sigma}-\bar{\varphi}$ are defined so that $\bar{\sigma}\phi\bar{\varphi}$ gives the plastic strain energy.

[14] T. R. Nelson, J. E. Dorn and E. R. Parker, "Investigations on the validity of an idealized theory of plastic flow for rolled mild steels," First Progress Report to ONR (January 1947).
[15] T. Ueda, *Sci. Rep.* Tohoku Imp. Univ. **19**, 473–498 (1930); **20**, 193–230 (1932).

change.[11] Consequently, the coefficients of plasticity change with large strains. Although it is possible to so formulate the stress-plastic strain rate equations that they can include this phenomenon, the calculations become tedious and unwieldy and the theory is not greatly advanced. Fortunately the anisotropic characteristics are fairly constant over rather large strains. For most problems of anisotropic plastic flow, therefore, it is appropriate to assume that the ratios of the coefficients of plasticity remain constant over nominal deformations.

Let

$$\dot{A}_{kl}=(\alpha_{kl}\dot{\phi}/\sigma), \tag{17}$$

where α_{kl} is a constant of anisotropy, and $\dot{\phi}$ is a generalized strain rate and σ a generalized stress which will be evaluated later. The equations of plasticity for anisotropic flow then become

$$\begin{aligned}\dot{\epsilon}_{xx}&=(\dot{\phi}/\sigma)\{\alpha_{11}\tau_{xx}+\alpha_{12}\tau_{yy}\},\\ \dot{\epsilon}_{yy}&=(\dot{\phi}/\sigma)\{\alpha_{12}\tau_{xx}+\alpha_{22}\tau_{yy}\},\\ \dot{\epsilon}_{xy}&=(\dot{\phi}/\sigma)\alpha_{44}\tau_{xy}.\end{aligned} \tag{18}$$

Since the α_{kl}'s are dimensionless ratios, they are not all independent, and it is therefore appropriate to place α_{11} equal to one. The following relationships are obtained for the various symmetries:

a. $x-z$ and $y-z$ plane: $\alpha_{11}=1$.
b. Fourfold z-axis: $\alpha_{11}=\alpha_{22}=1$; $\alpha_{21}=\alpha_{12}$.
c. Planar isotropy: $\alpha_{11}=\alpha_{22}=1$; $\alpha_{21}=\alpha_{12}$; $\alpha_{44}=2(1-\alpha_{12})$. (19)
d. Three dimensional isotropy: $\alpha_{11}=\alpha_{22}=1$; $\alpha_{12}=\alpha_{21}=-\frac{1}{2}$; $\alpha_{44}=3$.

WORK HARDENING FUNCTION

When metals work harden, the stresses necessary to cause plastic flow increase with deformation. It is reasonable to suspect that the increase in stress is some function of the plastic strain energy.[6,9] Since this assumption is known to be quite accurate for isotropic deformations of work hardenable metals, it will also be applied here. Then the definitions of σ and $\dot{\phi}$ will be completed by assuming that

$$\sigma=dW/d\phi, \tag{20}$$

where

$$dW=\sigma d\phi=(\tau_{xx}\dot{\epsilon}_{xx}+\tau_{yy}\dot{\epsilon}_{yy}+\tau_{xy}\dot{\epsilon}_{xy})dt. \tag{21}$$

Upon substitution of Eqs. (18)

$$dW=\sigma d\phi=\{\tau_{xx}[\alpha_{11}\tau_{xx}+\alpha_{12}\tau_{yy}]+\tau_{yy}[\alpha_{21}\tau_{xx}+\alpha_{22}\tau_{yy}]+\tau_{xy}\alpha_{44}\tau_{xy}\}(d\phi/\sigma). \tag{22}$$

Consequently,

$$\sigma=-\{\tau_{xx}(\alpha_{11}\tau_{xx}+\alpha_{12}\tau_{yy})+\tau_{yy}(\alpha_{21}\tau_{xx}+\alpha_{22}\tau_{yy})+\alpha_{44}\tau_{xy}^2\}^{\frac{1}{2}}. \tag{23}$$

Solving Eqs. (18) simultaneously for τ_{xx}, τ_{yy}, and τ_{xy}, and introducing these values into Eq. (21) yields

$$dW=\sigma\phi dt=\frac{\sigma}{\dot{\phi}}\left\{\frac{\dot{\epsilon}_{xx}[\alpha_{22}\dot{\epsilon}_{xx}-\alpha_{12}\dot{\epsilon}_{yy}]}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}+\frac{\dot{\epsilon}_{yy}[\alpha_{11}\dot{\epsilon}_{yy}-\alpha_{21}\dot{\epsilon}_{xx}]}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}+\frac{\dot{\epsilon}_{xy}^2}{\alpha_{44}}\right\}dt, \tag{24}$$

and, therefore,

$$\dot{\phi}=\frac{\partial\phi}{\partial t}=\left(\frac{\alpha_{22}\dot{\epsilon}_{xx}^2-(\alpha_{12}+\alpha_{21})\dot{\epsilon}_{xx}\dot{\epsilon}_{yy}+\alpha_{11}\dot{\epsilon}_{yy}^2}{\alpha_{11}\alpha_{22}-\alpha_{12}\alpha_{21}}+\frac{\dot{\epsilon}_{xy}^2}{\alpha_{44}}\right)^{\frac{1}{2}}, \tag{25}$$

where $\phi=\int(\partial\phi/\partial t)dt$ and the generalized flow stress is a function of the generalized strain and strain rate,

$$\sigma=\sigma(\phi,\dot{\phi}). \tag{26}$$

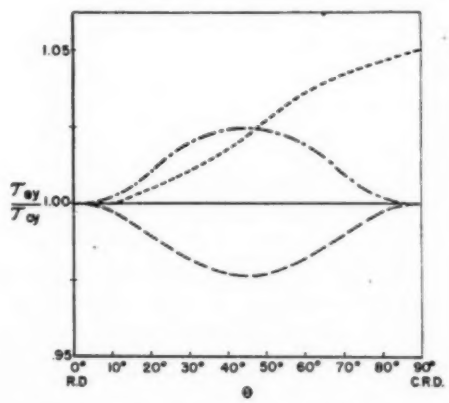


FIG. 2. Effect of orientation on the tensile yield strength.

Curve	Condition	α_{11}	α_{22}	α_{12}	α_{21}	α_{44}
1———	Three-Dimensional Isotropy	1.00	1.00	−0.50	−0.50	3.00
2———	Planar Isotropy	1.00	1.00	−0.45	−0.45	2.90
3———	Planar Isotropy	1.00	1.00	−0.55	−0.55	3.10
4— — —	Four-Fold Axis of Symmetry	1.00	1.00	−0.45	−0.45	3.10
5—·—·—	Four-Fold Axis of Symmetry	1.00	1.00	−0.55	−0.55	2.90
6--------	xz, yz, or xs and ys Plane of Symmetry	1.00	1.10	−0.40	−0.50	3.00

The values of σ and $\partial\phi/\partial t$ for higher types of symmetry are obtained by introducing the special conditions 19 into Eqs. (23) and (25). It is significant that for complete three-dimensional isotropy

$$\sigma=(\tau_{xx}^2-\tau_{xx}\tau_{yy}+\tau_{yy}^2+3\tau_{xy}^2)^{\frac{1}{2}}, \quad (27)$$

and

$$\phi=\int_0^t\frac{\partial\phi}{\partial t}dt$$

$$=\int\left(\frac{4}{3}(\dot{\epsilon}_{xx}^2+\dot{\epsilon}_{xx}\dot{\epsilon}_{yy}+\dot{\epsilon}_{yy}^2)+\frac{\dot{\epsilon}_{xy}^2}{3}\right)^{\frac{1}{2}}dt, \quad (28)$$

which are the commonly used equations for describing the work hardening characteristics of isotropic metals.[16]

Although many investigations have reported fair agreement of the above methods of correlation with experimental data on plastic flow of isotropic metals, only one investigation on anisotropic metals has provided the necessary data for checking the accuracy of the anisotropic theory.[2]

Tensile studies on a mild steel sheet revealed that it had planar isotropy where $\alpha_{12}=0.467$. Tension tests gave the same stress-plastic strain curves for all directions in the sheet. The work hardening curves ($\sigma-\phi$) for pure tension and equal biaxial tension are shown in Fig. 1. According to the above theory, the two methods of testing should have given identical $\sigma-\phi$ curves. The curves for simple tension and equal biaxial tension agree well at small strains, suggesting the nominal validity of the theory. A potential weakness of the theory, may possibly be inherent in the assumption of the linear relationship between the stresses and plastic strain rates. Prager[8] suggested that the strains might be an odd powered series of the components of the stresses in isotropic materials. In view of the general lack of experimental data for checking the present theory, it appears to be appropriate, at this time, to retain only the linear approximation.

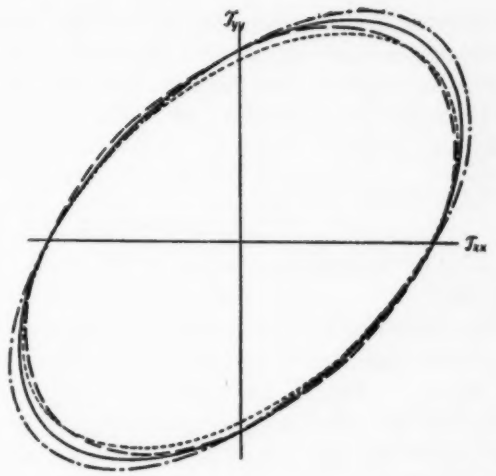


FIG. 3. Yield stress under combined stresses.

Curve	Condition	α_{11}	α_{22}	α_{12}	α_{21}
1———	Three Dimensional Isotropy	1.00	1.00	−0.50	−0.50
2— — —	Planar Isotropy or Four-Fold Axis of Symmetry	1.00	1.00	−0.45	−0.45
3— - - —	(Same as Curve 2)	1.00	1.00	−0.55	−0.55
4--------	xz, yz, or xz and yz Plane of Symmetry or Complete Anisotropy	1.00	1.10	−0.40	−0.60

YIELD STRENGTH

It will be useful to define the generalized yield strength as the flow stress, σ_0, when the plastic strain ϕ has reached a small specified value, $\phi=\phi_0$. The tensile yield strength for various orientations to the rolling direction are obtainable by introducing Eqs. (6) into Eq. (23) and letting $\tau_{x'x'}=\tau_{y'y'}=0$. Then,

$$\sigma_0=\tau_{y'y'}(\alpha_{11}\sin^4\theta+(\alpha_{12}+\alpha_{21}+\alpha_{44})\times\sin^2\theta\cos^2\theta+\alpha_{22}\cos^4\theta)^{\frac{1}{2}}. \quad (29)$$

The ratio of the tensile yield strength at an angle θ to the rolling direction ($\tau_{y'y'}=\tau_{y'\theta}$) to the yield strength in the rolling direction ($\tau_{y'y'}=\tau_{y'0}$) is

$$\frac{\tau_{y'\theta}}{\tau_{y'0}}=\left(\frac{\alpha_{22}}{\alpha_{11}\sin^4\theta+(\alpha_{12}+\alpha_{21}+\alpha_{44})\sin^2\theta\cos^2\theta+\alpha_{22}\cos^4\theta}\right)^{\frac{1}{2}}, \quad (30)$$

since σ_0 has the same value for yielding in the two tests.

Although there have been extensive studies on the effect of orientation on the tensile yield strength of sheet metals, none of the published studies contain the data necessary for a complete verification of Eq. (30). In general, however, the available evidence shows trends analogous to those recorded in Fig. 2 for various specializations.

The same procedure used to determine the

[16] A. Nadai, "Plastic behavior of metals in the strain hardening range," J. App. Phys. 8, 205–212 (1937).

conditions for yielding under simple tensile stressing can also be applied to yielding under combined stresses. For the simplest case where the principal stresses are in the rolling and cross-rolling directions

$$\sigma_0=[\tau_{xx}(\alpha_{11}\tau_{xx}+\alpha_{12}\tau_{yy})+\tau_{yy}(\alpha_{21}\tau_{xx}+\alpha_{22}\tau_{yy})]^{\frac{1}{2}}, \quad (31)$$

as given by substitution of $\sigma_0=\sigma$ and $\tau_{xy}=0$ into Eq. (23).

The combined stresses at which yielding occurs in various special cases of anisotropic plastic flow are shown in Fig. 3. The isotropic curve represented by the solid line is the well known Mises condition[17] for yielding. For anisotropic metals having planar symmetry or a four fold axis of symmetry, major and minor axes of the ellipse differ from the Mises condition. When the metal exhibits $y-z$, $x-z$, or $x-z$ and $y-z$ planes of symmetry or complete three-dimensional anisotropy, not only do the major and minor axes of the yield ellipse differ from the Mises ellipse, but also the ellipse is rotated from the Mises orientation of 45° to the directions of principal stress.

[17] R. von Mises, "Mechanik der festen Korper im plastisch-deformablem Zustande," Nadr. d. Gessellock. d. Wissensch. zu Gottingen, Math.-Phys. Klasse (1913).

ACKNOWLEDGMENT

The author wishes to express his appreciation for the suggestions and assistance given by the Office of Naval Research for conducting this investigation. He also thanks Mr. T. Hazlett for proof-reading the manuscript and Mrs. G. Pelatowski for preparing the figures.

The Theoretical Resolution Limit of the Electron Microscope

O. SCHERZER

Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey

(Received June 14, 1948)

The resolving power of the electron microscope and the contrast in the image are calculated for different conditions of focusing, illumination and aperture. These conditions can change the limit of resolution by a factor of about 3. The contrast in the image of an atom is appreciably increased by defocusing and spherical aberration. Nevertheless, the contrast improves when the numerical value of the aberration constant is diminished. The effect of different methods of spherical correction is discussed briefly.

ACCORDING to Abbe's theory the resolution limit of a light microscope with illumination parallel to the optic axis is given by

$$d=0.8\lambda/\sin\alpha, \quad (1)$$

where d is the minimum distance between two light-absorbing particles that can be recognized as separate in the image, λ is the wave-length in the object plane, and α is the aperture angle of the objective lens. If the aperture of the illumination is increased to equality with the objective aperture α, the resolution limit is improved to

$$d=0.6\lambda/\sin\alpha, \quad (2)$$

which also holds for self-luminous specimens.

In electron optics the relations between aperture and resolving power are complicated by the following facts:

(a) Electron lenses of the usual type cannot be corrected spherically. Electron waves passing through the outer zones of the lens miss the Gaussian image point and therefore do not contribute to resolution. The result is the same as though these outer zones were covered with an impenetrable screen. In many cases the resolving power is, therefore, not determined by the angular aperture α, but by the aberration constant C_A of the lens.

(b) In most cases only a small fraction of the observed contrast is caused by absorption of electrons in the specimen. The contrast is mainly caused by the phase delay of the electron waves on their way through the specimen. This phase delay may cause a deflection which makes some electrons hit the diaphragm or pass through the uncorrected outer zones of the lens, thus reducing the intensity of the image point; it may also cause some electron waves to interfere with less delayed electron waves, which gives rise to interference fringes in the image plane. Therefore, the formula giving the resolution for absorbing specimens cannot be used as a sufficient criterion for the visibility of an object unless the specimen is thick enough to cause detectable absorption.

(c) The scattering angles of the electrons are very small, so that the aperture of the objective is not filled unless the aperture is very small. When the diaphragm is not hit by electrons, it does not influence the resolving power and the contrast. This is another reason why a change of the aperture angle often does not influence the appearance of the image.

The recent improvement of the resolving power of the electron microscope[1] demands a careful investigation of the theoretical limitations of resolution and contrast. Therefore the equivalent of Eqs. (1) and (2) in the presence of spherical aberration will be derived here, and the contrast observed in the image of an atom will be calculated, using the scattering characteristics of the atom.

I. THE LIMIT OF RESOLUTION FOR ABSORBING PARTICLES

The first estimate of the influence of spherical aberration on the limit of resolution was given by the author in 1937.[2] Rebsch[3] and Dosse[4] improved the numerical calculations, and Glaser[5] gave a more adequate treatment of the problem by taking the wave nature of the electron fully into account. As all these results are restricted to images taken in the Gaussian image plane, a more general derivation will be given here.

Let us first assume the image plane to be at a very large distance from the objective lens. When we know the wave function ψ_a of the electron in a plane behind the lens, the wave function ψ_i in the image plane can be calculated according to Huygens's principle in Fraunhofer's approximation:

$$\psi_i(r_i, \vartheta_i) = \frac{1}{\lambda_i f_i M} \int_0^{r_m} r dr \int_0^{2\pi} d\vartheta \times \exp\left(2\pi i \frac{r_i \lambda \cos(\vartheta - \vartheta_i)}{M f_i \lambda_i}\right) \psi_a(r, \vartheta). \quad (3)$$

Here r, ϑ are the coordinates in the plane of integration. If the aperture diaphragm is arranged behind the lens, the r-coordinate r_m of the outermost ray is the radius of the aperture. M is the magnification, λ_i and f_i are wave-length and focal length in the image space; they are connected with the corresponding quantities λ and f in the object space by the relation

$$\lambda_i f_i = \lambda f.$$

An irrevelant complex phase factor of the type $i \exp(i\omega t - ikz)$ has been omitted in Eq. (3) and will be omitted in all following equations.

If the lens were ideal, the function $\psi_a(r, \vartheta)$ would be proportional to the scattering factor $S(\Theta, \vartheta)$ of the specimen, Θ being the scattering angle as shown in Fig. 1. When the spherical correction and the focusing of the lens are not perfect, the different zones of the lens cause different phase shifts of the electron waves. The phase shift γ as a function of r is defined by the equations

$$\psi_a(r, \vartheta) = (1/f) S(\Theta, \vartheta) e^{i\gamma}, \quad \Theta = (r/f),$$

and can be calculated in the following way.

In terms of geometrical optics the deviations of the ray angle from its ideal value zero are

$$\delta = \frac{C_A}{f_i}\Theta^3 - \frac{\Delta f}{f_i}\Theta = \frac{C_A r^3}{f_i f^3} - \frac{r\Delta f}{f_i f}. \quad (4)$$

C_A is the aberration constant, frequently denoted[5] by C_5 or[6] by $r_{aa}{}^3 S_8/M$ or Cf. The deviation of the focal length from the value f that would focus the paraxial rays in the image plane is Δf. In the

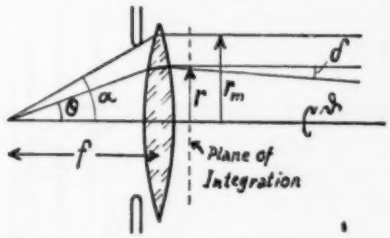


FIG. 1. The deviation δ caused by defocusing and spherical aberration.

TABLE I. Resolution factor $c_1 = (\lambda/C_A)^{\frac{1}{4}} d/\lambda$.

Aperture	Focus	c_1 for illumination: axial	optimal
∞	Gaussian	1.1	0.8
optimal	Gaussian	0.8	0.6
∞	optimal	0.7	0.5
optimal	optimal	0.6	0.4

[1] J. Hillier, J. App. Phys. **17**, 307 (1946).
[2] Quoted on page 559 of reference 3.
[3] R. Rebsch, Ann. d. Phys. **31**, 551 (1938).
[4] J. Dosse, Zeits. f. Physik **117**, 722, 1941.
[5] W. Glaser, Zeits. f. Physik **121**, 647, 1943.
[6] Zworykin, Morton, Ramberg, Hillier, and Vance, *Electron Optics and the Electron Microscope* (John Wiley & Sons, Inc., New York, 1945).

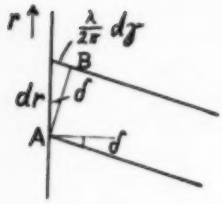


FIG. 2. Relation between deviation δ and phase shift γ.

wave description, the phase shift γ caused by the lens aberrations determines the wave surfaces in case of a point source object, for which $S(\Theta, \vartheta)=1$. As the rays are normal to the surface of constant phase (AB in Fig. 2), the angle δ is connected with γ by the differential equation

$$\delta=(\lambda_i/2\pi)(d\gamma/dr).$$

With the abbreviations

$$s=\left(\frac{\pi C_A}{2\lambda}\right)^{\frac{1}{4}}, \quad \tau=\frac{\pi\Delta f}{\lambda s}=\left(\frac{2\pi}{\lambda C_A}\right)^{\frac{1}{4}}\Delta f, \qquad (5)$$

the solution may be written

$$\gamma=\frac{2\pi}{\lambda_i}\int\delta dr=\frac{s^2r^4}{f^4}-\frac{s\tau r^2}{f^2}=s^2\Theta^4-\tau s\Theta^2.$$

If the scattering function does not depend on the azimuth ϑ, the integral over ϑ is the Bessel function of order zero:

$$\psi_i(r_i)=\frac{2\pi}{M\lambda}\int_0^\alpha \exp[i(s^2\Theta^4-\tau s\Theta^2] \times S(\Theta)J_0\left(\frac{2\pi r_i}{M\lambda}\Theta\right)\Theta d\Theta. \qquad (6)$$

When the object under observation is a small hole in an opaque foil, the scattering function $S(\Theta)$ is practically constant within the angle α, so that $S(\Theta)$ can be put equal to its value at $\Theta=0$. The remaining integral has been evaluated carefully by Glaser[5] for the special case $\tau=0$, i.e., for the case that the image is taken in the Gaussian image plane. For the more general case $\tau\neq 0$ we shall use a less rigorous method, which nevertheless gives the right values of the resolving power within the uncertainty caused by the somewhat arbitrary mathematical definitions of this quantity. For this purpose we notice that the exponential function in (6) varies slowly for small Θ and oscillates rapidly for larger Θ. For not too large values of r_i the Bessel function varies slowly with Θ, so that the regions with rapidly oscillating exponential function do not contribute appreciably to the integral. As the transition between the two modes of behavior of the exponential function is rather abrupt (Figs. 3 and 4), we replace the function by a constant in the region where it varies slowly, and by zero in the region of rapid oscillation. The integral (6) can then be evaluated analytically:

$$\psi_i(r_i)=\frac{\Theta_m}{r_i}S(0)J_1\left(\frac{2\pi r_i}{M\lambda}\Theta_m\right), \qquad \Theta_m=\begin{cases}\alpha & \text{if } \alpha<\Theta_0\\ \Theta_0 & \text{if } \alpha>\Theta_0\end{cases}, \qquad (7)$$

where Θ_0 is the value of Θ at which the exponential function begins to oscillate.[7]

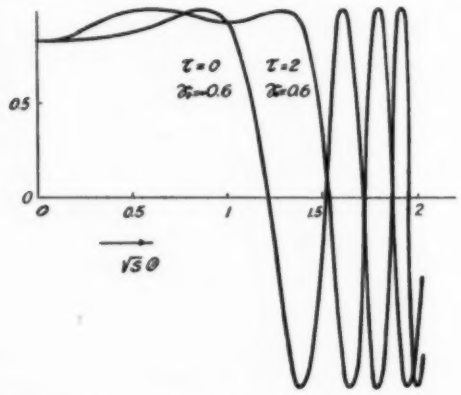


FIG. 3. The function $\cos(s^2\Theta^4-\tau s\Theta^2+\gamma_0)=Re\{e^{i\gamma_0}e^{i\gamma}\}$, describing the phase shift caused by defocusing and spherical aberration.

If the object consists of two small holes separated by a distance d, and the illumination is axial (i.e., parallel to the axis), the ψ_i of both holes coincide in phase and have to be added. The distance d is equal to the resolving power, when a just perceivable minimum occurs in the sum curve.

The mathematical definition of a just perceivable minimum is somewhat arbitrary. In accordance with Fig. 19.9 of reference[6] we consider a minimum as just perceivable when its intensity $I_{\min}$ is 75 percent of the maximum $I_{\max}$. In the more general case that the image appears

[7] Where the results were sensitive to this value, Θ_m was defined as the lowest positive solution of the equation

$$Re\left\{\int_{\Theta_m}^{\alpha} e^{i\gamma}\Theta d\Theta \Big/ \int_0^{\Theta_m} e^{i\gamma}\Theta d\Theta\right\}=0.$$

on a background of intensity I_∞, we have to take the contrast $g=1-I/I_\infty$ into account. The condition then becomes

$$I_{max}-I_{min}=(I_{max}-I_\infty)/4,$$

regardless of whether the differences are positive or negative. This definition is in accordance with Eqs. (1) and (2).

If the illumination is oblique under an angle α_0, the ψ_i differ in phase by $2\pi\alpha_0 d/\lambda$. When the condensor aperture α_C is about λ/d, all phase differences occur, and the ψ_i have to be added incoherently. Thus the resolution limit d is characterized by the appearance of a minimum in the sum of the $|\psi_i|^2$.

When the resolving power for not too small apertures is determined in this way, the effective limiting angle Θ_m in (7) becomes proportional to $s^{-\frac{1}{4}}$, where s is given by Eq. (5), and the result can be written in the form

$$d=c_1\lambda(C_A/\lambda)^{\frac{1}{4}}. \qquad (8)$$

The resolution factor c_1 depends on the values of τ and $\alpha s^{\frac{1}{4}}$. Some typical values can be taken from Table I, in which optimal focusing signifies the focusing for which c_1 is smallest. This occurs for $\tau\approx 2.5$, which means that the inner zones of the lens are a little too weak for exact (Gaussian) focusing, characterized by $\tau=0$. The optimal aperture is $\alpha=1.2(\lambda/C_A)^{\frac{1}{4}}$ for $\tau=0$, and $1.4(\lambda/C_A)^{\frac{1}{4}}$ for $\tau=2.5$. Aperture ∞ means α large compared with these values.

If the object consists of two absorbing particles on an ideally transparent foil, the calculations are complicated by the radiation which bypasses the particles and causes the scattering function $S(\Theta)$ to have a sharp peak at $\Theta=0$. This peak gives rise to the bright background in the image plane, on which the particles appear as dark spots. For axial illumination the phase of the background is constant and opposite to that of the scattered wave. The contributions of the two particles are in phase and the resolving power is the same as for two holes. For oblique illumination the phase of the background varies and the contributions of the two particles differ in phase. The calculation is rather involved, but within the uncertainty of the definition of resolving power the results are identical with Table I.

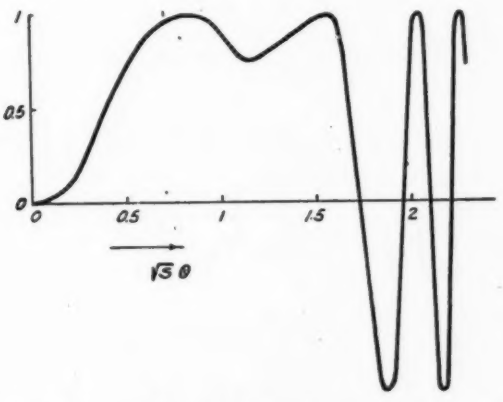


FIG. 4. The function $\sin(3s\Theta^2-s^2\Theta^4)$, describing the phase shift in case of optimum contrast.

TABLE II. Contrast of atoms with $Z>5$ in case of axial illumination.

Range of d	Range of α	Focus	Contrast g
$Z^{1/3}d>12$A	$C_A\alpha^4\ll\lambda$	Gaussian	$0.6Z^{4/3}\lambda^2/d^2$
$1.2\text{A}<Z^{1/3}d<11$A	$C_A\alpha^4\ll\lambda$	Gaussian	$1.3\cdot10^{-3}\beta^{-1}Z^{5/3}\lambda/d$
$Z^{1/3}d<1$A	$C_A\alpha^4\ll\lambda$	Gaussian	$3\cdot10^{-5}Z^2/\beta^2$
$d>5$A	$C_A\alpha^4\geq\lambda$	Gaussian	$-1.2\cdot10^2\beta Z^{2/3}\lambda^2/d^3$
$d>5$A	$C_A\alpha^4\geq\lambda$	optimal	$1.7\cdot10^2\beta Z^{2/3}\lambda^2/d^3$
$1\text{A}<Z^{1/3}d<12$A	$C_A\alpha^4\geq\lambda$	optimal	$1.2Z^{2/3}\lambda/d$
$Z^{1/3}d<0.6$A	$C_A\alpha^4\geq\lambda$	optimal	$0.04Z/\beta$

Rebsch has shown,[3] that in the best of the practically attainable systems C_A is about $0.25f$. For such systems we may write

$$d=0.7c_1\lambda(f/\lambda)^{\frac{1}{4}}. \qquad (9)$$

When we use (8) or (9) to derive the values of c_1 from the values of d given in references 2 to 5, we find 1.4, 1, 1.32, 0.56, respectively. This checks relatively well with the values given in Table I for different conditions. The experimental values of d found by Hillier[1] yield values of c_1 in the same order of magnitude.

We have derived Eq. (6) under the assumption that Θ is small and M is large. When Θ is relatively large, as in light optics, the only change necessary in (6) is to replace Θ by $\sin\Theta$, provided that the optical system fulfils Abbe's sine condition $r=f\sin\Theta$. As Θ dropped out of our results, Table I also holds for such systems. When M is not large compared with $\Theta_m^2 f/\lambda$, the plane of integration in Fig. 1 has to be replaced by a sphere around the center of the image plane. The mathematical procedure is complicated by this fact, but the results are not changed. This can be understood from the fact that mathematically

the optical system can be split up in a non-ideal system producing a virtual image at infinity, and an aberration-free system projecting this virtual image into the real image plane without changing the relative intensities. Equation (6) and Table I can therefore be used quite generally when the angular aperture is of the order $(\lambda/C_A)^{\frac{1}{4}}$ or larger.

II. THE CONTRAST IN THE IMAGE OF AN ATOM

The resolving power given by Eq. (8) and Table I is only one of the two quantities that determine the visibility of an object. The other important quantity is the contrast. When the absorption in the specimen is relatively high, the contrast depends mainly on the area over which the image of an object element is spread in the image plane, i.e., on the resolving power of the microscope. Yet in many specimens in electron microscopy the absorption is negligible. When there is no absorption at all, an ideally imaging microscope would not show any contrast at all, because all the scattered electrons would be gathered in the image point, making its intensity equal to that of the background.

For an aberration-free lens with finite aperture and image in the Gaussian image plane, the contrast g caused by single atoms can be calculated[8,9] from the equation

$$g=[I(\infty)-I(0)]/I(\infty)=\sigma/d^2. \qquad (10)$$

$I(r_i)$ is the intensity in the image plane, $r_i=0$ being the center of the image; σ is the effective cross section of the atom for scattering through an angle larger than the angular aperture α, and d^2 is the area in the object plane, over which the atom seems to be blurred because of the finite resolution limit d. Equation (10) follows from the assumption that the intensity loss caused by electrons hitting the diaphragm is spread uniformly over the whole image disk of the atom, and also holds when more accurate assumptions about the intensity distribution are made.[8]

The first three lines of Table II give the results for the Thomas-Fermi model of the atom. They have been calculated by using Boersch's approximations for small and for large scattering angles and by approximating Boersch's curve (reference 9, p. 91) by a straight line in a part of the intermediate region. The symbol β has its usual meaning $\beta=v/c=1\text{A}/41\lambda$.

Spherical aberration and deviations from the Gaussian focusing cause an additional phase shift of the electron waves which adds to the phase shift caused by the specimen and makes non-absorbing objects visible even if no electrons are caught by the aperture-limiting diaphragm. To calculate the contrast in such cases we have to know the scattering function $S(\Theta)$ in Eq. (6). $S(\Theta)/r$ is the wave function on a sphere of large radius r around the scattering atom. If the wave-length is short compared with the dimensions of the atom, as is the case for electrons above 30 kv, the phase delay γ_a of an electron going in z-direction through the potential field $\varphi=\varphi(r)$ of an atom is given by

$$\gamma_a=\int(k-k_0)dz; \quad k=\frac{2\pi}{\lambda}=\frac{2\pi mv_0}{h}\left(1+\frac{2e\varphi}{mv_0^2}\right)^{\frac{1}{2}}.$$

Since in most parts of the atom the potential energy $e\varphi$ is small compared with the kinetic energy, the square root can be expanded and yields

$$k-k_0\approx\frac{2\pi e\varphi}{hv_0}; \quad \gamma_a=\frac{2\pi e}{hv_0}\int\varphi dz.$$

Using again Huygens's principle as in Eq. (3), and writing x, y for the Cartesian coordinates of the element of integration in a plane within or just behind the atom, we find the wave function

$$\frac{1}{r}S(\Theta)=\frac{1}{\lambda r}\int\int\exp(ik\Theta y)e^{-i\gamma_a}dxdy.$$

The smallness of γ_a allows the second exponential function to be replaced by $1-i\gamma_a$:

$$S(\Theta)=\frac{1}{\lambda}\int\int\exp(ik\Theta y)dxdy$$

$$-\frac{2\pi ie}{\lambda hv_0}\int\int\int\exp(ik\Theta y)\varphi dxdydz. \qquad (11)$$

When the atom is in the middle of a diaphragm of radius r_0, the first integral can be evaluated

[8] See reference 6, p. 729.

[9] H. Boersch, Monats. f. Chemie **76**, 86 and 163 (1946). Note added after completion of this paper: Recently a more complete treatment has been given: H. Boersch, ZS. f. Naturforschung **2a**, 615 (1947).

in polar coordinates in a way similar to the step from (3) to (7), yielding

$$\frac{1}{\lambda}\int\int \exp(ik\Theta y)dxdy=\frac{r_0}{\Theta}J_1(kr_0\Theta). \quad (12)$$

The second integral may be simplified in the usual way[10] by writing

$$\exp(ik\Theta y)=-\frac{1}{k^2\Theta^2}\Delta\exp(ik\Theta y).$$

Integrations by parts make the Laplacean operator Δ operate on the potential φ. Since $\Delta\varphi$ is -4π times the charge, and the charge is the charge Ze in the nucleus minus the electronic charge distributed with the density $e\rho$, we find

$$\int\int\int \exp(ik\Theta y)\varphi dxdydz$$
$$=\frac{-1}{k^2\Theta^2}\int\int\int \exp(ik\Theta y)\Delta\varphi dxdydz$$
$$=4\pi e\frac{Z-F(\Theta)}{k^2\Theta^2}.$$

Here $F(\Theta)$ is the usual atomic scattering factor

$$F(\Theta)=\int\int\int \rho\exp(ik\Theta y)dxdydz, \quad (13)$$

which also describes the scattering of x-rays and therefore has been investigated very carefully. To find the wave function in the image plane, the scattering function (11) has to be inserted into Eq. (6). As r_0 is very large, the part (12) contributes only for $\Theta=0$, so that the other functions in (6) can be taken for $\Theta=0$, and the only variable in the integrand is J_1, the integral of which is $-J_0$. Thus we have

$$M\psi_i(r_i)=1-\frac{4\pi ie^2}{hv_0}\int_0^\alpha \exp[i(s^2\Theta^4-s\tau\Theta^2)]$$
$$\times(Z-F)J_0\left(\frac{2\pi r_i}{M\lambda}\Theta\right)\frac{d\Theta}{\Theta}. \quad (14)$$

As the intensity I is proportional to $|\psi_i|^2$ and the second term on the right side is a small number, we find for the contrast g defined by (10):

$$g=\frac{2\pi e^2}{hv_0}\int_0^\alpha (Z-F)\sin(s\tau\Theta^2-s^2\Theta^4)\Theta^{-1}d\Theta. \quad (15)$$

In this approximation the contrast vanishes for $s=0$, $s\tau=0$. For finite values of s the integral can be evaluated by using the known values of $F(\Theta)$. For $Z>5$ and not too small values of Θ the Thomas-Fermi model can be used. The F-values for this model[11] can be approximated by the formulae

$$Z-F=60Z^{\frac{1}{3}}\beta\Theta, \quad \text{if} \quad 6\cdot10^{-4}Z^{\frac{1}{3}}<\beta\Theta<1.5\cdot10^{-2}Z^{\frac{1}{3}},$$
$$Z-F=Z, \quad \text{if} \quad \beta\Theta>0.02Z^{\frac{1}{3}}.$$

At very small angles the Thomas-Fermi model cannot give the right values, because the scattering function in this region is determined by the electron distribution in the outer parts of the atom, where the electron density is so small that corpuscular statistics cannot be applied. The strong increase of the function $\mathfrak{F}'$ in reference 11 at very small angles is caused by this faulty behavior of the Thomas-Fermi model. The true scattering function in this region depends on the valence electrons and cannot be described by a power of Z. Nevertheless a sufficient approximation for our purposes is obtained by setting the scattering function equal to the value of the Thomas-Fermi scattering function at $\Theta=\lambda/6A$, because at this angle the electrons within a sphere of 6A diameter contribute primarily, while the electrons outside this sphere, which are there only in the model but not in the real atom, do not. We therefore use the approximation

$$\frac{Z-F}{\Theta^2}=\left(\frac{Z-F}{\Theta^2}\right)_{\Theta=\lambda/6A}=2\cdot10^4Z^{\frac{1}{3}}\beta^2,$$
$$\text{if} \quad \beta\Theta<6\cdot10^{-3}.$$

With these approximations the integral (15) can be evaluated with relative ease for different values of s, τ, and α. For instance for $\tau=0$ and $\alpha=\infty$ the formula

$$\int_0^\infty x^n\sin x^4dx=\frac{1}{1+n}\sin\frac{\pi(1+n)}{8}\Gamma\left(\frac{5+n}{4}\right)$$

may be used. In other cases, e.g., for $\tau=3$ as shown in Fig. 4, the sine function in (15) can be

[10] N. F. Mott, Proc. Roy. Soc. **A127**, 658 (1930).

[11] A. Sommerfeld und H. Bethe, Handbuch der Physik **24**, 2, 423, Fig. 34 with $\xi=\Theta/(2\lambda Z^{\frac{1}{3}})$.

considered constant in a wide region and zero outside. Thus the last four lines of Table II have been obtained. The symbol d denotes the resolving power given by Table I for axial illumination and the corresponding focusing and aperture conditions. Focusing for optimal contrast means $\tau \approx 3$, while optimal resolution is obtained with $\tau \approx 2.5$. Yet the resolving power is not very sensitive to an increase of τ, and the values given for optimal focus in Tables I and II hold also for $\tau = 3$ within the limits of error.

When the illumination aperture is comparable with the objective aperture α, the formulae for the contrast are rather complicated. It can be expected, however, that the contrast is not greater than given by Table II. In most cases of wide angle illumination it is much less.

With $C_A = 0.8$ cm and $\lambda = 0.05$A, Table I, gives a resolving power of 4A under optimal conditions, and 6 to 11A under different conditions of axial illumination. In an aberration-free microscope of this resolving power even the heaviest atoms would be invisible in Gaussian focus, because for $C_A = 0$, $Z = 92$ and $d = 6$A, Table II yields $g = 0.017$, which is at least a factor of 2 below the threshold of observation.[9] For $C_A = 0.8$ cm and Gaussian focus the heaviest atoms should be visible ($d = 8$A, $Z = 90$, $g = -0.05$). However, the contrast is negative in the center of the image. The atoms appear as bright spots of diameter Md within dark rings of radius Md. In the optimal focus all atoms heavier than chromium should show up with positive contrast ($d = 6$A, $Z = 90$, $g = 0.12$; $Z = 24$, $g = 0.05$), appearing as dark spots within bright rings of radius Md.

With a resolving power of 2A in axial illumination, a value which cannot be achieved with microscopes of present day type, the heaviest atoms should be visible without the help of aberrations and defocusing ($\lambda = 0.05$A, $Z = 90$, $g = 0.12$). For optimal focusing Table II yields $g = 0.6$ for heavy atoms and $g = 0.10$ for $Z = 6$, so that even single carbon atoms would be visible. A molecule like C_6I_6 with the relatively large distances between the iodine atoms would surely show its hexagonal symmetry.

Table II further shows that for values of d far below 1A the contrast is independent of the resolving power, as has already been stated by Hillier.[12] The result found for such small values of d and optimal focus resembles a formula derived by Schiff[13] for $Z^{\frac{1}{3}}d \approx 3$A, which gives the right order of magnitude. Yet it should be kept in mind that for $d \ll 1$A the loss of contrast caused by the heat motion of the atoms cannot be neglected.

For bare nuclei the last line of Table II holds without restrictions of Z and d. If the charge of a proton were not compensated by an electron or by the polarization of its environment, it would be near the threshold of observation independently of the resolving power of the microscope.

As has been pointed out by Schiff,[13] the contrast can be increased by observation in dark field. Yet dark field usually deteriorates the resolving power because of the very inhomogeneous distribution of the scattered electron waves over the aperture of the lens. The contrast of bright field images is increased by inelastic scattering,[9] but only within the limits of error of our calculations. The supporting film distorts the axial illumination and thus decreases the contrast. This is one more reason to make the supporting foils as light as possible.

The results check with the known fact that not all steps which improve the resolving power also improve the contrast, and that a small amount of aberrations and defocusing is beneficial to the contrast. But they also show clearly that an improvement of the resolving power, gained by a decrease of the spherical aberration, improves the contrast in the image of an atom as long as the limit of resolution is above 1A.

III. CONTRAST AND RESOLUTION OF LARGER OBJECTS

When the specimen is not too thick, multiple scattering of electrons can be neglected and the total scattered wave is merely the sum of the scattered waves from the single atoms. If the diameter of the particle under observation is smaller than $d/2$, the scattered waves are practically in phase. If all atoms have the same atomic number Z, the contrast given by Table II is to be multiplied by the number N of the atoms in the particle. In Gaussian focus the particles

[12] J. Hillier, Phys. Rev. **60**, 743 (1941).
[13] L. I. Schiff, Phys. Rev. **61**, 721 (1942).

will appear as bright spots, surrounded by dark rings of radius Md. In the focusing which gives optimal contrast they appear as dark spots within bright rings of the same size. In intermediate focus, i.e., when spherical aberration and defocusing partly compensate each other, the interference fringes and the contrast are relatively weak. The contrast is then about N times the contrast of an atom imaged by an aberration-free lens in Gaussian focus, given by the first three lines of Table II.

When the atoms are distributed uniformly over a large area, as in a foil, and the resolution limit d is large compared with the separation of the atoms, the bright and the dark parts of the interference fringes overlap and compensate each other, so that only the absorption of electrons in the specimen and in the diaphragm give rise to contrast. When the diaphragm limits the resolving power and there are n equal atoms per unit area of the foil, the number of particles per object element is approximately nd^2, and the contrast is nd^2 times the value of g from the first three lines of Table II. If there is no aperture diaphragm, the contrast in the image of homogeneous foils is due to the absorption in the foil only and is much less than in microscopes with an aperture diaphragm, as is known from experiments.

Near the edge of a foil the compensation of dark and bright regions of the interference fringes is not complete. In Gaussian focus we have an overbalance of the bright spots near the edge and an overbalance of the dark rings outside the geometrical image of the foil. In the focus which gives optimum contrast the intensities are reversed. The distance between the dark and the bright lines is approximately the radius of the first ring in the image of an atom, i.e., Md.

The rings around particles as well as the fringes along edges of foils can be seen very clearly in electron micrographs.[14] As long as the limit of resolution is large compared with the distance of the atoms, these contour effects can also be understood by considering the foil as a homogeneous medium of known refractive index.[14]

[14] J. Hillier and E. G. Ramberg, J. App. Phys. **18**, 48 (1947), Fig. 22. E. Kinder und A. Recknagel, Optik **2**, 346 (1947), Fig. 2.

When the resolution limit is determined experimentally by measuring the minimum distance between resolved particles, regions with regularly arranged particles should be avoided. Particles arranged on the foil in a regular way act like a two- or three-dimensional diffraction lattice. They may cause the scattered electron waves to form a small number of well defined beams. If the most important of these beams go through the same zone of the lens, the spherical aberration does not interfere with the sharpness of the image, and an unexpectedly good resolution may be obtained. Of course this resolution does not characterize the microscope but only the crystalline or quasi-crystalline arrangement of the test particles. If two isolated particles or two particles out of a random distribution are chosen, such errors do not occur.

IV. THE INFLUENCE OF SPHERICAL CORRECTION

Our results show that in many cases of electron microscopy the contrast as well as the resolving power depend on the spherical aberration of the objective lens. Unfortunately the condition of spherical correction is in mathematical contradiction to the simultaneous validity of the following three conditions which characterize electron microscopes of present day type:

1. The imaging electric and magnetic fields are axially symmetric.
2. The fields are free from sources and vortices in the neighborhood of the optic axis.
3. The imaging fields do not vary with time.

As soon as one of these three conditions is abandoned the lenses can be corrected spherically and chromatically.[15]

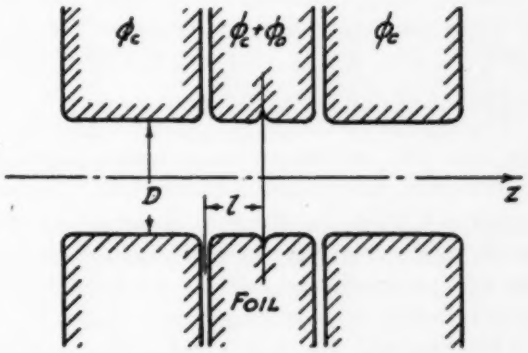


FIG. 5. Spherical correction by a foil.

[15] O. Scherzer, Optik **2**, 114 (1947).

An example is given by a lens system containing a conductive foil on which electrostatic charges are influenced by neighboring electrodes. If the electrodes have proper shape, the influenced charges will compensate the spherical aberration of the other parts of the system. A correcting lens of this type is shown in Fig. 5. The two outer cylinders of diameter D have the potential ϕ_c against the cathode. The inner cylinder of length $2l$ and the foil in its central plane $z=0$ have the potential $\phi_c+\phi_0$. The influence of the foil on the potential can be found by the principle of imaging. Equation [13.50] of reference 6 then gives the approximate potential on the axis:

$$\phi(z)-\phi_c=\phi_0\left(\frac{1}{2}\tanh\frac{2.63(l-|z|)}{D}-\frac{1}{2}\tanh\frac{2.63(l+|z|)}{D}\right).$$

If ϕ_0 is small compared with ϕ_c, the deviation δ of a ray entering this system parallel to the axis is

$$\delta=-\frac{r}{2}\frac{\phi'_0}{\phi_c}+\frac{r^3}{16}\frac{\phi'''_0}{\phi_c}, \qquad (16)$$

where ϕ'_0 and ϕ'''_0 are the derivatives of ϕ for $z=0$ on the right side of the foil:

$$\phi'_0=-2.63(1-t^2)\phi_0/D,$$
$$\phi'''_0=-2\cdot 2.63^3(4t^2-3t^4-1)\phi_0/D^3,$$
$$t=\tanh(2.63l/D).$$

It is seen that for $D=2l$, $\phi_0=\phi_c/8$ and $D=0.45f\cdot(f/C_A)^{\frac{1}{4}}$ the second term in (16) just compensates C_A in (4) for $f=f_i$, while the first term can easily be compensated by a slight change in Δf. The correcting system, therefore, compensates the spherical aberration when arranged in or behind an uncorrected lens.

The choice $l=0.25D$ makes ϕ'''_0 vanish. In this case the correcting system itself is a lens free from third-order spherical aberration. The focal length is $1.15D\phi_c/\phi_0$. Because of our assumption $\phi_0\ll\phi_c$ this is relatively long. Therefore, the system will preferably be used as a correcting element in addition to other lenses.

Unfortunately the foil is not only a support for the correcting surface charges, but it also acts on the electron waves by the scattering power of its atoms. When there are n atoms per unit area of the foil and the total scattering cross section of each atom is σ, the resolving power as well as the contrast will suffer when $n\sigma>0.5$.

When $n\sigma<0.2$, multiple scattering can be neglected and the intensity distribution in the image plane can be calculated from Eq. (6) after adding the phase delay in the atoms of the foil to the phase delay in the lens described by the exponential function. As the scattered waves have 90° phase difference to the unscattered wave, the intensity distribution is not changed considerably in the first approximation. In the second approximation the coherent part of the background as well as the wave scattered by the object are changed in amplitude by a factor of $(1-n\sigma)^{\frac{1}{2}}$. The contrast, as the product of both, is therefore decreased by a factor of $1-n\sigma$. The resolving power is not affected as long as $n\sigma\ll 1$.

The usual collodion foils do not fulfil the condition $n\sigma\ll 1$ very well. Beryllium foils can be prepared much thinner.[16] With a thickness of 15A and the value of σ given by Marton and Schiff,[17] we find $n\sigma=0.16$. Hence a beryllium foil of this thickness should not interfere with resolving power and contrast and should allow the correction of microscopes

Other methods of spherical correction avoid the foil, but have other disadvantages. A cloud of free electrons[18] or a wire stretched along the optic axis[19] can be used as carriers of the correcting charges. But it is difficult to stabilize the electron cloud, and the wire cuts the middle part out of the lens. Spherical correction by the image force, whose field has a finite divergence and draws the electrons away from the axis, can be used for very slow electrons only. For voltages above 10 kv the long narrow channel necessary for this type of correction does not allow a sufficiently large image field and makes the alignment practically impossible.

Giving up the axial symmetry[15] or using

[16] V. E. Cosslett, Biochimica et Biophysica Acta 2, 239 (1948). I thank Dr. V. E. Cosslett for the privilege of seeing the manuscript before publication.

[17] L. Marton and L. I. Schiff, J. App. Phys. 12, 759 (1941).

[18] D. Gabor, *The Electron Microscope* (Hulton Press, London, 1946).

[19] D. Gabor, Nature 158, 198 (1946); J. W. Dungey and Catherine R. Hull, Proc. Phys. Soc. 59, 828 (1947).

alternating fields[20] also leads to spherical correction, but aligning of axially asymmetric lenses or working with ultra-high frequency will not be simple either. Nevertheless, there is a well founded hope that the spherical aberration will be overcome and a higher resolving power will be reached.

When the third-order aberration can be varied at will, it is better not to make it exactly zero, but to choose a value of C_A that will match best with the value C_5 of the fifth-order aberration constant. Then we have instead of Eq. (8)

$$d=0.6c_1\lambda(C_5/\lambda)^{1/6}, \qquad (17)$$

with c_1 given by Table I and the contrast given again by Table II. As C_5 will be of the order of magnitude of the focal length f like the constant C_A of uncorrected systems, the resolution limit d is lower than that of uncorrected systems by a factor of about 8.

[20] See reference 6 p. 632, and reference 15 Fig. 7. In the latter unfortunately 25 kv has been printed instead of 2.5 kv.

In the system shown in Fig. 5 the ratio of fifth-order to third-order aberration depends on the diameter D. If D differs on the two sides of the foil, a change in the ratio of the voltages applied on the two sides will change the ratio of the different orders. Thus the third-order and the fifth-order aberrations can be varied independently, and fifth-order correction will be possible. This means that the sixth root has to be replaced by an eighth root and the resolving power will be improved by another factor of 2. About the same factor also holds for the other correcting procedures. It is, therefore, to be expected that single atoms will be visible as soon as one of the methods of spherical correction proves to be practically feasible.

ACKNOWLEDGMENT

I wish to express my thanks to the members of the Micro-Optical Section, Components and Materials Branch, Signal Corps Engineering Laboratories, for many helpful discussions.

The Internal Friction of Zinc Single Crystals

CHARLES A. WERT

State University of Iowa, Iowa City, Iowa

(Received June 14, 1948)

This paper presents typical data on the internal friction of twelve single crystals of zinc oscillating longitudinally. The measurements of decrement were made at frequencies of 41 kc and 57 kc. (a) Zinc of three grades of purity was used. The behavior of decrement as a function of stress amplitude for these different materials was found to depend less on the purity of the metal than has been reported previously (see references 1 and 2). (b) For one crystal the decrement showed little rise from its initial value of 2×10^{-5} at very low stress amplitudes up to stress amplitudes nearly three times the static elastic limit. Considerable similarity of behavior of the decrement as a function of stress amplitude resolved into the basal plane has been found for a number of crystals with a wide range of orientation. The obtaining of these consistent results depended on the use of appropriate annealing and etching techniques. (c) The decrement of a well annealed crystal increased with rising temperature at all stress amplitudes throughout the temperature range −75°C to 150°C. (d) The prior application of static stress far below the static elastic limit has been found to increase the decrement at high stress amplitudes without affecting that at very low stress amplitudes. This increase was a function of the stress amplitude at which the decrement was measured, and of the prior static stress, L. The observations fitted a relation of the form $\log\Delta=C_1+C_2L$, where C_1 depended on the amplitude of measurement and C_2 was approximately constant.

INTRODUCTION

THE internal friction of zinc single crystals has been measured previously by several investigators.[1–3] The most striking result of all of these measurements is that internal friction is greatly influenced by the treatment of the crystal prior to the measurement. Indeed, internal friction may well be the most sensitive indicator of a state of strain in a crystal. Read[1]

[1] T. A. Read, Phys. Rev. **58**, 371 (1940).

[2] T. A. Read and E. P. T. Tyndall, J. App. Phys. **17**, 713 (1946).

[3] I. H. Swift and J. E. Richardson, J. App. Phys. **18**, 417 (1947).

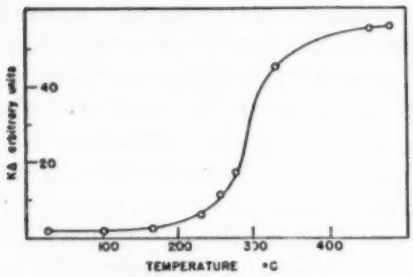


FIG. 1. The change in the product $K\Delta$ with temperature for the 41-kc quartz crystal.

first used crystals made of extremely pure zinc.[4] By annealing these crystals appropriately, he was able to reduce the internal friction measured at low strain amplitudes (less than 10^{-7}) to a value comparable to that of quartz crystal. However, the internal friction rose very rapidly with increasing strain amplitude. Working with zinc of slightly less purity,[5] Read[2] found the internal friction to be about as low as for the SP zinc at low strain amplitudes. Furthermore, it remained low with increasing strain amplitude, even, in some cases, to well beyond the static elastic limit. Swift and Richardson,[3] also using BH zinc, were unable to obtain such low values for the internal friction of their crystals at low stress amplitudes. The behavior at increasing stress amplitudes, however, agreed, in general, with Read's work on this same zinc. The crystals of Swift and Richardson were of much smaller cross section than those used by Read, and this enhanced the difficulty of handling the crystals without producing accidental strains. For this reason, even though they measured more crystals than had previously been done, their results are not without certain ambiguities, particularly in regard to the value of internal friction measured directly after growth and to the efficacy of subsequent annealing. In fact, more than half the crystals grown by them had internal friction too high to be measured in the apparatus, and no amount of annealing reduced it. There was no obvious reason for this, since all crystals were supposedly grown and handled alike.

[4] Read's analysis shows that his zinc was 99.999 percent pure. While this zinc and the purest zinc used in the present work do not have identical impurity content, both will be designated spectrographically pure (SP).

[5] This zinc was Bunker Hill, a commercial zinc whose total impurity content is 0.01 percent or less. This zinc will be designated BH.

The present investigation started with several objects in view. The first of these consisted in obtaining consistent and reproducible results on a set of crystals of larger cross section grown from BH zinc, studying, as before, the dependence of internal friction on the amplitude of vibration, the effect of etching, annealing, etc. Also, new measurements were made on SP crystals in the same apparatus under the same conditions, to find what differences might be caused by the slight difference in purity. In addition, internal friction was studied as a function of temperature for both zinc and as a function of previously applied static stresses. Read[2] has shown that previous application of static stress far below the elastic limit is sufficient to cause a marked change in the internal friction of a well annealed crystal.

EXPERIMENTAL PROCEDURE

The apparatus and methods of measurement were the same as those used by Swift and Richardson.[3] Two quartz crystals were used, resonant at 41 kc and 57 kc, respectively. These crystals were square in cross section, with an area of 0.25 sq. cm. The measurement of the decrement of a specimen depends on a prior knowledge of the electromechanical constant, K, and the decrement, Δ,[6] of the quartz. The constants, K and Δ, for each of the crystals were measured in a manner described by Cook.[7] The decrement of each crystal was reduced from 50×10^{-5} to 1.5×10^{-5} by etching it in dilute hydrofluoric acid. Because use was to be made of the quartz above room temperature, it seemed wise to make some check on any possible variation of K and Δ with temperature. The product of the two is easily measured, and a plot is shown in Fig. 1. Up to about 150°C, there is little change from the room temperature value. Since the variation of K with temperature is probably less than 10 percent in this temperature range,[8] for the purpose of calculating the decre-

[6] The decrement, a measure of the internal friction, is the ratio of the energy dissipated per cycle to twice the total vibrational energy. It is related to Q by the relation $\Delta=\pi/Q$.

[7] W. T. Cook, Phys. Rev. **50**, 1158 (1936).

[8] The constant K depends on certain piezoelectric and other constants of quartz, and the experimental evidence of their variation with temperature is not too exact, but seems to be as stated above. At higher temperatures the

ments of the zinc crystals it was assumed that K was constant and that the slight rise in the curve was due to a change in Δ only. Measurements of decrements of zinc crystals were not made above 150°C, and at this temperature the use of the value of K obtained at 25°C would not introduce an error of more than 10 percent in the measured decrements of the zinc crystals. For measurements to be made above 150°C, the behavior of the quartz would have to be more carefully investigated. Down to −75°C the product $K\Delta$ appeared to be constant at the room temperature value, and here too the room temperature determination of K was used to calculate the decrement of the zinc.

During measurement, the quartz and zinc were supported on two fine filaments of glass placed at displacement nodes of vibration, the quartz crystal being placed between two parallel brass plates which served as electrodes. These brass plates were mounted on a piece of transite, and the assembly was placed in the middle of a long glass tube. Electrical connections were made by running leads through rubber stoppers in the ends of the tube. The assembly was heated above room temperature by means of an electric furnace which was slipped over the tube; cooling below room temperature was accomplished by packing solid CO_2 around the tube. Temperatures were measured with a Chromel-Alumel thermocouple placed a few mm from the specimen. To reduce air damping, all measurements were made at a pressure of 0.1 mm of Hg or less.

The crystals used in the present investigation were grown from two grades of zinc—four from spectrographically pure zinc, ten from Bunker Hill zinc,[9] and one from BH, with the addition of 0.16 percent commercial copper. At the conclusion of the work, a spectrographic analysis was made of a number of the crystals and of both original materials. This analysis showed that the original SP zinc contained, besides lead, detectable amounts (estimated to be distinctly less than 0.001 percent) of copper and cadmium. No iron lines could be identified as such in the spectrum of the SP zinc. The original BH zinc contained about the same amount of copper and distinctly more cadmium and lead. In addition, three sensitive lines of iron and one of silver were detected for this zinc. No difference in purity was detected between the original materials and the crystals used in the investigation. As a check on the spectrographic analysis. Mr. S. W. Rabideau, of the University Chemistry Department, ran a polarographic analysis for lead on both original zincs. This analysis showed lead to be present in a barely detectable amount (surely less than 0.001 percent) for both, with no positive indication of difference between them.

The four crystals of SP zinc and some of the BH crystals were grown in Pyrex glass tubing coated with Aquadag. The remainder of the BH crystals were grown in transite molds by a method described by Cinnamon.[10] After growth the crystals were cut to one-half wave-length by

TABLE I. Composition and orientation of the crystals.

Crystals grown from Bunker Hill zinc			
B1	$\theta=90°$	B5	$\theta=63°$
B2	$\theta=82°$	B6	$\theta=39°$
B3	$\theta=75°$	B7	$\theta=20°$
B4	$\theta=75°$		
BC	BH+0.16% Cu		$\theta=58.5°$
Crystals grown from SP zinc			
S1	$\theta=63°$	S3	$\theta=39°$
S2	$\theta=56°$		

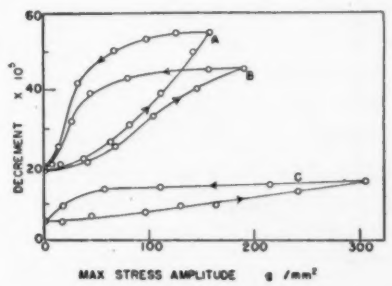


FIG. 2. For a partially annealed crystal, an anneal for 6 days at room temperature reduced the high stress amplitude decrement for curve A to curve B without any lowering of the low stress decrement. A subsequent anneal at 350°C lowered the decrement at all stress amplitudes. These curves were made for crystal B1 ($\theta=90°$) at 57 kc.

experimental evidence is somewhat conflicting. For a discussion of this point see W. Cady, *Piezoelectricity* (McGraw-Hill Book Company, Inc., New York, 1946), p. 223.

[9] The SP zinc was a gift from the New Jersey Zinc Company. This zinc is well known to contain less than 0.001 percent impurity. From information furnished by the New Jersey Zinc Company, the principle impurity in this lot was 0.0004 percent Pb. The BH zinc was from the same lot (50 No. slab) as that used by Read and Tyndall (see reference 2).

[10] C. A. Cinnamon, Rev. Sci. Inst. 5, 187 (1934).

means of a jewelers' saw, after being waxed into a piece of channel iron. Great care was taken to saw the ends of the crystal flat enough so that the ends needed no further grinding. The ends of the crystal were then etched in dilute HCl to remove the burns and cold work caused by the sawing operations.

Previous investigators have used beeswax and rosin and other materials to cement the zinc to the quartz. The writer found two other materials to be more suitable. These were phenyl salicylate, which was particularly easy to use at room temperature, and Insalute cement, usable over the entire temperature range.

RESULTS

Of the fifteen crystals grown, twelve were found to be suitable for use, and these were all used during the investigation. In this way any result on a single specimen was verified by making similar measurements on one or more other crystals. All these data are not given in detail, but typical data have been chosen for presentation here. The crystals for which data are given will be designated by the symbols given in Table I. The decrements of the twelve crystals measured at low stress amplitude varied from 2×10^{-5} to 10×10^{-5}. There was no apparent dependence on the purity of the zinc nor on the orientation[11] of the crystals. The decrement of the remaining three crystals was higher than 100×10^{-5}, and no amount of annealing and etching reduced it.

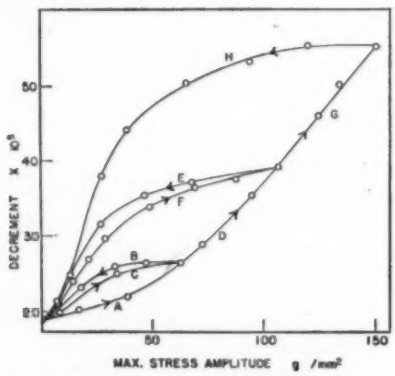


FIG. 3. Internal friction of crystal B1 as a function of increasing and decreasing amplitudes of vibration. The sequence of the measurements is shown by the letters and the arrows on the curves.

[11] The orientation is defined as the angle between the normal to the basal plane and the specimen length.

SURFACE EFFECTS

The loss of energy through surface effects can be considerable. For this reason the decrement of a well annealed crystal can be reduced to a still lower value through etching. Removing as little as 0.1 percent of the zinc by etching in dilute HCl may reduce the decrement 50 percent or more. The effect of removal of the surface layer was to lower the internal friction curve as a whole without changing the dependence of the decrement on the stress amplitude. To lower the low stress decrement to 2×10^{-5}, a number of anneals and etchings were, in general, necessary. The crystals grown in glass were found to require less etching than the ones grown in transite, since they were smoother and had a thinner layer of oxide than the latter; however, the lowest values of decrement obtainable did not depend on the method by which the crystals were grown.

ANNEALING

For all crystals which were mounted without a previous high temperature anneal, holding at room temperature after mounting resulted in a decrease in the decrement. For a typical case, the decrease in the decrement at low stress amplitudes was rapid for the first three days. Thereafter, it decreased more gradually so that after a month no further decrease was evident. Curve *A*, Fig. 2, is a typical plot of decrement *vs.* stress amplitude taken after such an anneal. Curve *B* was taken 6 days later, showing that even though the low amplitude decrement was constant, annealing was still taking place. Further annealing at room temperature produced no appreciable change in curve *B*. However, annealing the crystal for 12 hours at 350°C reduced the decrement greatly, as shown by curve *C*.

The rate of annealing is highly dependent on temperature. This is true for the decrement at all values of stress amplitude. Observations on 6 crystals showed that an anneal for a few hours at 350°C is as effective in reducing the decrement as several weeks at 180°C. For crystals left to rest for as long as 5 months at room temperature, an anneal at 350°C reduced the decrement still further. However, once a crystal was well annealed, moderate strains (i.e., strains which

increased the decrement at low stress to about 25×10^{-5}) were found to disappear rather quickly, even at room temperature.

Strains introduced by the measurement itself have been found to disappear rapidly, the original curves being retraceable in two days or less. In fact, recovery begins immediately after the measurement. Data taken for crystal B1 are shown in Fig. 3. The order in which the data were taken is shown in the figure, curves *A*, *B*, *C*, etc., following in alphabetical order, the arrows indicating increasing or decreasing stress amplitude between the points on the curves. The fact that curves *C* and *F* lie under curves *B* and *E*, respectively, shows that appreciable recovery had already taken place in the rather short time interval between the "down" and "up" curves. From these data it is not possible to determine whether this recovery is caused by annealing or by the unloading of the specimen.

PURITY

Read[1,2] has reported that there is considerable difference in the decrements at large stress amplitudes between crystals grown from SP and BH zinc. The present work agrees with his measurements on BH zinc, but indicates much less difference between the two than he has reported. Some of the results obtained in the present work, and replots[12] of curves made by Read for SP zinc, are shown in Fig. 4. There is striking difference between Read's curves *B* and *C* and the writer's curves S1 and S2, although they are made for crystals of roughly the same orientation. This difference can possibly be accounted for in one of the following ways: (a) The difference in nature of the impurity in the SP zinc used in the two investigations is an important factor, or (b) the crystals used in the present investigation were more free of strains. In regard to the first possibility, it may be said that that though the two materials apparently contain about the same percentage of impurity, the actual impurities are different metals. Read's zinc contained Fe as the principal impurity, with smaller concentrations of Pb and Cd; the writer's zinc contained Pb as the principal impurity, with lesser amounts of Cu and Cd. The writer

[12] These plots were estimated from his published curves, reference 1.

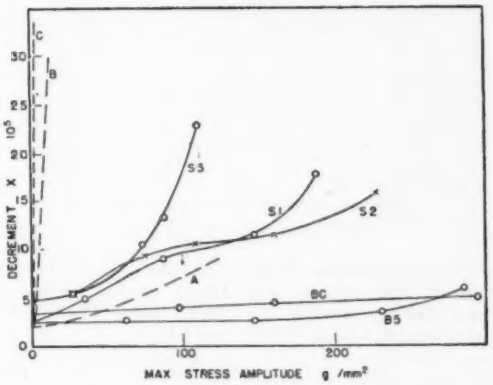


FIG. 4. The internal friction as a function of stress amplitude for five crystals used in the present work and three crystals used by Read.[1] Curves *A*, *B*, and *C* refer to Read's spectrographic crystals $\theta=88°$, $\theta=74.5°$, and $\theta=61.3°$, respectively. The rest of the curves refer to crystals described in Table I.

believes it to be possible that this difference in purity may alone be responsible for the difference in the data obtained in the two investigations. However, mention ought also be made of the second point above. Read[1] has reported that the low stress decrement for at least one of his crystals soon after mounting was about 500 $\times10^{-5}$. In the present investigation, the decrements of the SP crystals immediately after mounting did not exceed 75×10^{-5}, after which sufficiently long anneals at room temperature reduced the decrements to the values in curves. S1 and S2. It seems possible that with the original high decrements of Read's crystals, the anneals which he gave them may not have been sufficient to relieve all the strains in the crystals. It is true that he was able to reduce the low stress decrement to low values, but numerous measurements made in the present investigation show that this alone is not sufficient to ensure that the decrement will remain low with increasing stress amplitude. The writer does not suggest that these two possibilities are the only two explanations of the differences in the data, nor that one is more plausible than the other.

TEMPERATURE

In handling crystals there is always a possibility of producing strains, particularly for orientation between 25° and 65°. Therefore, it seemed that a very desirable technique would be to anneal the crystal after attachment to the

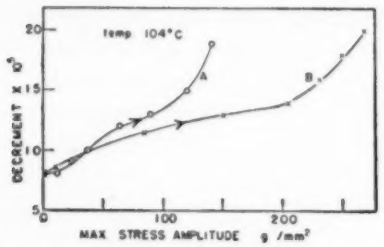


FIG. 5. An anneal for 11 days at 104°C was effective in eliminating the strains introduced at the quartz-zinc interface during cooling at 104°C from 350°C. These data are for crystal B2 ($\theta=82°$) $f=41$ kc.

quartz and while mounted in the apparatus. This would exclude any handling whatsoever after annealing. This procedure was found to be successful for crystals of low orientation ($\theta<20°$), but to produce poor results for higher orientations. For an 82° crystal the effect found is illustrated by Fig. 5. Curve *A* was taken at 104°C after the crystal had been cemented with Insalute, kept at 60°C for 6 hours to set the cement, raised to 350° and kept there for 12 hours, and then cooled to 104°C for the measurement. The crystal was then left at 104°C for 11 days, at which time curve *B* was run. After the anneal at 350°C, the subsequent lowering of decrement found after the 104°C anneal was certainly not to be expected. No such lowering occurred for low orientation crystals treated in the same way, nor for high orientation crystals ($\theta>65°$) when the anneal at 350°C preceded cementing and mounting. The explanation of the observed effect seems to lie in the fact that for any crystal of orientation greater than 20° the transverse thermal expansion of the zinc mismatches considerably the expansion of the quartz in one direction in one quartz-zinc interface, although in the direction perpendicular to this the match is good. For low orientation crystals, the match to the quartz is fairly good for all transverse expansions.[13] Thus the zinc crystal (Fig. 5) was strained while being cooled from 350° to 104°C, and, in spite of the fact that the strained portion was in a position of small effect on the decrement (being near a stress and strain minimum), it appeared to act in a deleterious fashion.

The mechanism which produces internal friction is possibly the same as that responsible for plasticity; hence it is reasonable to expect that the decrement would increase with rising temperature, since the ease of plastic deformation is known to increase with increase of temperature. The behavior of the decrement at a number of temperatures is shown for crystal B7 in Fig. 6A. For this crystal the decrement at all stress amplitudes does increase at higher temperatures. However, attempts to make such measurements on some other crystals met with certain difficulties which were eventually traced to the just described thermal expansion effect. The results of measurements made on crystal B2 are shown in Fig. 6B. The expected variation with temperature occurred at the lowest stress amplitudes, but at higher stress amplitudes the curve for −76°C was anomalous. The explanation for this behavior is found in the consideration of the large strains introduced in this crystal during cooling, coupled with the fact that very little annealing could take place at this low temperature. Similar strains occur on heating, but at the higher temperatures annealing takes place rapidly, so that these strains are relieved and these curves are more nearly normal. For crystal BC the results in Fig. 6D show it to behave in a more nearly normal fashion. The copper in this crystal had a hardening effect, so that the lowest temperature curve no longer shows the sudden rise above the other curves. The mismatching of thermal expansion for this crystal of 58.5° orientation is not as bad as for the 82° crystal (Fig. 6B), but analysis of the thermal stress actually gives a greater resolved shear stress in the slip plane, so that, except for the hardening effect, this crystal might well be even less nearly normal than the 82° crystal. This is shown rather strikingly by crystal S2, Fig. 6C, which has about the same orientation as BC. Here the results are very far from normal. This difference in behavior of these two crystals is apparently caused solely by the difference in purity of the metals.

[13] For quartz the coefficients of thermal expansion parallel and perpendicular to the "*C*" axis are 9.0×10^{-6} and 14.8×10^{-6} per degree centigrade, respectively. For zinc the corresponding coefficients are 64.1×10^{-6} and 14.1×10^{-6}, respectively; for zinc the coefficients in a given direction are given by $\alpha_\theta=(14.1+50\cos^2\theta)\times10^{-6}$, where θ is the angle between the given direction and the *C* axis.

COLD WORK

The effect on internal friction of previously applied static stresses has been studied for three crystals. Two methods of application of the external stresses were investigated. The first of these consisted of application of longitudinal stress by weights while the crystal was suspended with its axis vertical. Small brass buttons were cemented to the crystal to apply the load evenly across the end. A number of attempts to use this method gave results which were too uncertain to be reliable—primarily because it involved too much handling with the attendant accidental strains. The second method consisted of the application of a transverse force at the center of the crystal while it was being supported horizontally on a blunt knife edge and roller near its ends. It is believed that with the rather small forces involved at the points of support and application of the force, little damage results at the surface of the crystal. The stresses produced in the crystal by the bending moment are not uniform, but this disadvantage is more than offset by the fact that few manipulations are required, eliminating accidental strains almost completely. Data for crystal B5 are shown in Fig. 7. Curve *A* shows the decrement for the crystal in its initial well annealed state. The

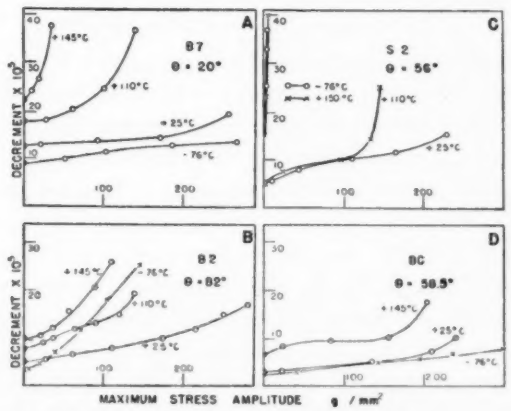


FIG. 6. The decrement, Δ, as a function of stress amplitude at a number of temperatures measured for four crystals. What is believed to be the normal behavior of the decrement at these temperatures is shown in A. The anomalous effect for badly mismatched crystals is shown in B and C. The addition of a small amount of Cu hardens the zinc so that the decrement is not so adversely affected by the differential expansion at the quartz-zinc interface, curves D. With the assumption that K is constant in the temperature range used, these plots may be in error at −76°C and +150°C by ±10 percent.

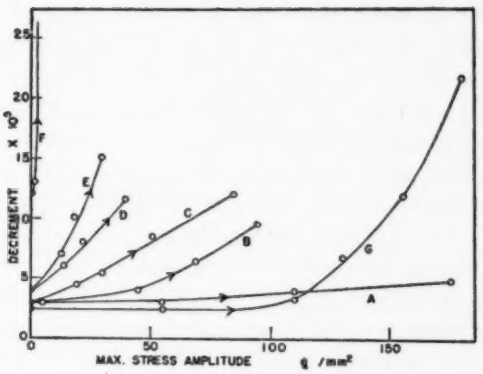


FIG. 7. The increase in internal friction of crystal B5 produced by successive static stresses of 25, 50, 75, 100, and 150 g/mm², each applied for five minutes, together with the decrease caused by a subsequent anneal for two weeks at room temperature.

crystal was then removed from the apparatus, set on the end supports, and a transverse force applied at the middle such that the maximum longitudinal stress in the zinc was 25 g/mm². After five minutes the force was removed, the crystal remounted in the apparatus, and the data shown in curve *B* were taken. Curves *C*, *D*, and *E* followed in the same fashion with the maximum stresses being 50, 75, and 100 g/mm², respectively. When a load of 150 g was applied, the crystal was left noticeably permanently bent after removal of the load. Curve *F* followed this bending. The crystal was then straightened and allowed to rest for two weeks at room temperature, after which curve *G* was taken. These measurements show that static loads considerably below the static elastic limit[14] (for this crystal about 65 g/mm²) cause an appreciable increase in internal friction at high stress amplitudes. This agrees qualitatively with a previous measurement on this zinc by Read.[2] Furthermore, the decrement measured at low stress amplitude did not increase appreciably until the crystal was grossly deformed, after which time the recovery was rapid. Data taken in a similar fashion for crystals B2 and S1 gave results in agreement with those of Fig. 7.

An analysis of the curves in Fig. 7 shows that the decrement at constant stress amplitude of measurement, σ, is related to the previously applied static stress by the relation

$$\Delta = A10^{b\sigma},$$

[14] This term is described in some detail in reference 2.

where A and b are constant for a particular σ, and L is the magnitude of the previously applied static stress. This is shown in Fig. 8 for crystal B5 by curves A, B, and C. Except for the points of zero applied static stress, the data fit the equation quite well. The slopes, b, of the curves lie between 0.0185 and 0.0225, with an average value of 0.021 mm²/g. Similar data for crystal S2 are shown by curves D, E, and F, with slopes from 0.0185 to 0.029, averaging 0.023 mm²/g. The data are not very extensive but seem to indicate a rather definite relation, particularly since the slopes are roughly the same.

DISCUSSION

As has been mentioned previously, the mere attainment of a low decrement at low stress amplitudes does not necessarily indicate complete freedom from strain in a crystal. This is shown rather strikingly by the static stress experiments described in a preceding paragraph. Even after the crystal lattice was greatly deformed by permanent bending of the crystal, the low stress decrement increased to only 12×10^{-5}; but the rapid rise of decrement with increasing stress amplitude showed that the crystal was severely strained. Further bending apparently would have resulted in a further rise of the low stress decrement. This does not mean, however, that any crystal which has a relatively high decrement at low stress will necessarily have a decrement which will increase rapidly as the stress is increased. It is possible that a crystal may have a low stress decrement of 50×10^{-5} or higher, and that the decrement will not increase markedly from this value until the stress amplitude rises above the static elastic limit. This is amply demonstrated by Swift and Richardson,[3] and by additional data obtained in the present research.

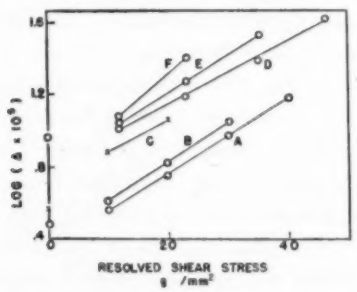


FIG. 8. A plot of logΔ *vs.* the static shear stress of a previously applied load resolved into the basal plane. Curves A, B, and C correspond to stresses of measurement of 12, 16, and 32 g/mm² *RSS*, respectively, for crystal B5; curves D, E, and F correspond to 5.5, 7, and 9.5 g/mm² *RSS*, respectively, for crystal S2.

If internal friction is caused solely by surface effects and lattice imperfections, it would seem that by careful etching and annealing one ought to be able to reduce the low stress decrement to an arbitrarily low value, but such is not the case; the lower limit for zinc is about 2×10^{-5} at 25°C. A decrement of this magnitude is, however, nearly as low as that of crystal quartz, so its existence is not surprising. Whether this residual decrement is caused by lattice imperfections, surface effects, or some other mechanism is not apparent at present, but there is reason to believe that it is independent of the amplitude of the applied stress. This has been discussed by Swift and Richardson,[3] they propose that the decrement at a given stress is composed of two parts, one part the residual decrement and the other the energy loss caused by movement of dislocations. This point of view is assumed in the paragraph to follow.

There is a different significance attached to the behavior of the decrement as a function of stress amplitude than to that measured at low stress amplitude. There are, however, some difficulties in comparing data taken at high stress amplitudes, one of which is the dependence of the data on the crystal orientation. Comparison may be more meaningful if the decrement be plotted as a function of the stress resolved into the basal plane. This is perhaps justified by the belief that internal friction is caused at least in part by movement of dislocations in the basal planes. Data for nine of the crystals used in the present work, and for two others reported previously,[2,3] are shown in Fig. 9. The abscissa in this plot is the resolved shear stress and the ordinate is the decrement at a given stress less the decrement at low stress (call this $\Delta_S=\Delta-\Delta_0$). Four of the curves (A, B, $B5$, $B7$) lie lower than the rest, (B, $B5$ and $B7$) for the entire stress range and A for the first half). For them, Δ_S does not begin to increase rapidly until the stress amplitude is higher than the static elastic limit;[15] for three of

[15] For BH zinc the static elastic limit, as determined by Tyndall (private communication), is a resolved shear stress of about 25 g/mm².

them (*B*, *B*5, and *B*7) until the stress amplitude is nearly three times the static elastic limit. The remaining seven crystals are quite alike to stresses up to the elastic limit, after which the curves diverge somewhat. All these crystals were cemented with Insalute cement, and, with the exception of *S*2, they were all annealed alike. That variation in annealing is important is shown by the difference in curve *S*2 and the other six. Data for curve *S*2 were taken following an anneal for 8 weeks at room temperature after mounting. The flattening out of curve *S*2 at high stress amplitudes is not surprising, since it has been shown that a long room temperature anneal is effective in flattening out a curve of Δ *vs.* stress amplitude.

Several conclusions may be drawn from these data. It is believed that for a strain free crystal curves *B*, *B*5, and *B*7 represent the true behavior of the decrement as a function of resolved shear stress. With proper treatment the decrement of any crystal of zinc (comparable to the present material) should be expected to behave similarly. That curve *B*7 is in this group, although it too was cemented with Insalute, is due to the fact that for this crystal the previously mentioned thermal strain was absent, so that high temperature anneal after mounting was very effective. Crystal *B*5 was allowed to anneal for a long time at room temperature after mounting, using phenyl salicylate as cement, so it too was free from strains. Curves *S*1, *S*3, *B*2, *B*3, *B*4, and *B*6 cannot be considered as representing crystals which were entirely strain free. Presumably, for these crystals not enough time was allowed after mounting for the thermal strains to disappear.

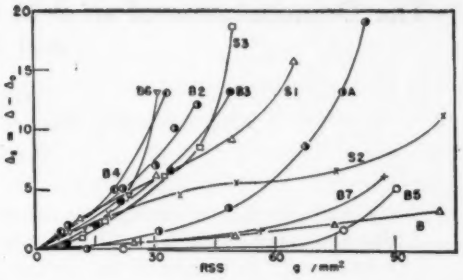


FIG. 9. A plot of Δs as a function of the maximum resolved shear stress for a number of BH crystals used in the present work, and two BH crystals used in previous investigations. These last two are (a) Read, $\theta = 70°$, curve *A*; and (b) Swift and Richardson, $\theta = 70°$, curve *B*.

They do show, however, that, using a particular technique, consistent results may be obtained for a large group of crystals with a rather wide range of orientation. Finally, comparison of curves *A*, *B*, *B*5, and *B*7 show the difficulty of attempting to correlate the results of different investigators even when they use crystals made of the same grade of zinc. It appears that while the results of a single investigator using a particular procedure may form a consistent set of data, direct comparison of the results of various investigators is made less significant by the variation in experimental details. The decrement is apparently so sensitive to strains and surface effects that differences in growing, cutting to length, handling, cementing, and annealing a crystal may give it a history which it is not possible to eliminate completely.

In conclusion, the author wishes to express his gratitude to Professor E. P. T. Tyndall for his interest and advice throughout the course of the investigation.

Calendar of Meetings

January

19–21 American Society of Civil Engineers, New York, New York
24–27 American Society of Heating and Ventilating Engineers, Chicago, Illinois
24–27 Institute of Aeronautical Sciences, New York, New York
27–29 American Physical Society and American Association of Physics Teachers, New York, New York
31–February 4 American Institute of Electrical Engineers, New York, New York

February

3–5 American Physical Society, Berkeley, California
13–17 American Institute of Mining and Metallurgical Engineers, San Francisco, California
21–24 Technical Association of the Pulp and Paper Industry, New York, New York
26 American Mathematical Society, New York, New York, and Chicago, Illinois
28–March 4 American Society for Testing Materials (Spring Meeting), Chicago, Illinois

March

3–4 Illuminating Engineering Society, San Francisco, California
6–10 American Institute of Chemical Engineers, Los Angeles, California
7–10 Institute of Radio Engineering, New York, New York
8–10 Society of Automotive Engineers, Detroit, Michigan
9 Inter-Society Color Council, New York, New York
10–12 American Physical Society, Cleveland, Ohio

The Demodulation of an F-M Carrier and Random Noise by a Limiter and Discriminator*

NELSON M. BLACHMAN
*Cruft Laboratory, Harvard University, Cambridge, Massachusetts***
(Received June 17, 1948)

Certain mathematical representations of a limiter and discriminator are postulated. On this basis, and on the basis of the statistics of narrow-band random noise, the output signal voltage and r.m.s. noise output voltage are calculated for a limiter and discriminator fed by an f-m carrier and narrow-band random noise. The output signal-to-noise ratio can thus be expressed as a function of the input signal-to-noise ratio. The former is plotted *versus* the latter in Fig. 5, for several different degrees of limiting. A sharp increase in the output signal-to-noise ratio from far below the input signal-to-noise ratio to only 0.9 db below it is found to occur as the latter increases through the range 0 db to 6 db. For the sake of comparison, Fig. 5 also shows the corresponding curve for square law detection of an a-m carrier, as well as that for f-m without a limiter. It is to be emphasized that these results pertain to the unfiltered discriminator output; consideration of filtering, which is generally present in a receiver to a considerable extent, requires a determination of the spectral distribution of the discriminator output noise.

INTRODUCTION

ABOUT thirty years ago considerable interest was aroused by the idea that frequency modulation might afford a means of compressing a signal into a much narrower band of frequencies than that required for amplitude modulation. However, in 1922 Carson showed[1] that f-m required at least as wide a band as a-m. Interest waned somewhat until very high frequencies came into use, since at these frequencies band width is not a particularly restrictive factor. In 1936 Armstrong showed[2] in a qualitative manner that wide-band f-m has certain inherent noise- and interference-reducing properties not possessed by a-m. "Wide-band" f-m is that case in which the maximum change in the carrier frequency is several times the highest audio modulating frequency and a correspondingly wide band is required for the transmission of the signal (modulated carrier). When the maximum change in the carrier frequency is comparable with or smaller than the highest audio modulating frequency, the band width requirements are much like those of a-m; this case is known as "narrow-band" f-m.

Following Armstrong's paper, several mathematical demonstrations of the ability of f-m to discriminate against noise and interference under appropriate circumstances appeared,[3–6] most notably those of Carson and Fry.[7,8] All of these demonstrations consider only signals which are much stronger than the interfering noise voltages. With such signals, wide-band f-m definitely can give a better signal-to-noise ratio than a-m, provided a limiter is used.

The functional relation between the audio output signal-to-noise ratio and the carrier-to-noise ratio is known for a-m receivers.[9–12] It

* This article summarizes a part of the author's doctoral dissertation, *The Demodulation of a Frequency-Modulated Carrier and Random Noise by an F-M Receiver*, which appeared as Cruft Laboratory Technical Report No. 31 (March 5, 1948). The work was supported successively under O.S.R.D. Contract OEMsr-1441 and joint Office of Naval Research-Signal Corps Contract N5ori-76, T. O. I, with Cruft Laboratory.

** Now at the Brookhaven National Laboratory, Upton, Long Island, New York.

[1] J. R. Carson, Proc. I.R.E. **10**, 57 (1922). See also B. Van der Pol, Proc. I.R.E. **18**, 1194 (1930).

[2] E. H. Armstrong, Proc. I.R.E. **24**, 689 (1936).

[3] H. Roder, Electronics **10**, 60 (May, 1937).

[4] M. G. Crosby, Proc. I.R.E. **25**, 472 (1937).

[5] E. H. Plump, Hoch:tech u. Elek:akus. **52**, 73 (1938).

[6] O. E. Keall, W. Eng. **18**, 6, 56 (1941). This paper concerns the interference of two signals.

[7] J. R. Carson and T. C. Fry, Bell Sys. Tech. J. **16**, 513 (1937).

[8] J. R. Carson, Bell Sys. Tech. J. **18**, 395 (1939).

[9] R. E. Burgess, *The Rectification of Signal and Noise by Linear and Square-Law Detectors* (Radio Research Board Report C93, England, March 24, 1944).

[10] S. O. Rice, Bell Sys. Tech. J. **24**, 46 (1945), continued from **23**, 283 (1944).

[11] J. R. Ragazzini, Proc. I.R.E. **30**, 277 (1942).

[12] D. Middleton, "Some general results in the theory of noise through non-linear devices," Quart. App. Math. **5**, 445 (1948). For signal-to-noise ratio following general half-wave detection, see "Rectification of a sinusoidally modulated carrier in the presence of noise," paper accepted for publication in Proc. I.R.E.

remains to determine the corresponding result for f-m receivers in order to be able to determine whether the carrier competes more effectively with the noise in an a-m or an f-m system. It is assumed that the noise is of the random, fluctuation type, i.e., hissy, and that there is no impulse noise.[13] (Recently Rice extended[14] his work[10] to cover some f-m aspects of random noise, especially for the case of extreme limiting; Middleton has extended[15] the present work and found the spectral distribution of the discriminator output noise, which is needed in order to be able to determine the amount of *audible* noise output.)

The essential distinction of an f-m receiver lies in its using a limiter and discriminator in place of a detector. We shall, therefore, consider the action of a limiter and discriminator on an f-m carrier and noise. For this purpose we need analytic representations of the two devices.

ANALYTIC REPRESENTATION OF LIMITER

Since the input to the limiter has passed through the narrow-band (compared to the intermediate frequency) i-f amplifier, it consists of sine waves of nearly equal frequency and, therefore, resembles a sine wave of the intermediate frequency. It may, in fact, be considered to be a phase- and amplitude-modulated sine wave of the intermediate frequency. The purpose of the limiter is to remove all of the fluctuations in amplitude. The output of the limiter should preserve the phase of its input, but the output amplitude should be a constant, say V_1.

However, there is a certain minimum input amplitude, V_0, for which the output amplitude is still V_1. When the input amplitude, V, exceeds V_0, the output amplitude is V_1. When V is less than V_0, the limiter is presumably operating at its maximum gain, and so the output amplitude is proportional to the input amplitude, i.e., it is $(V_1/V_0)V$. Letting $V_1 = V_0$ and allowing them both to become infinite corresponds to the case in which no limiter is used.

[13] See V. D. Landon, Electronics **14**, 37 (Aug., 1941).
[14] S. O. Rice, Bell Sys. Tech. J. **27**, 109 (1948).
[15] D. Middleton, *The Spectrum of a Frequency-Modulated Wave after Reception in Random Noise. I* (Tech. Rep. No. 33, Cruft Laboratory, Harvard University) (March 8, 1948). *On Theoretical Signal-to-Noise Ratios in F-M Receivers: A Comparison with Amplitude Modulation* (Cruft Report No. 38) (June 18, 1948).

ANALYTIC REPRESENTATION OF DISCRIMINATOR

The instantaneous output voltage of a discriminator should be proportional to the deviation of the instantaneous frequency of its input from the intermediate frequency ω_0, since the output of the discriminator is then a reproduction of the signal modulating the transmitter. This response to the instantaneous frequency corresponds to the "quasi-stationary" hypothesis of Carson and Fry.[7]

The output voltage of a discriminator using linear rectifiers is necessarily proportional to the amplitude V' of its input voltage, for a fixed input frequency. Combining this fact with the assumption that the output voltage is proportional to the deviation D' (from the intermediate frequency) of the input voltage, we obtain the complete expression for the discriminator output voltage,

$$kV'D',$$

in which k is a constant of the device measured in the unit seconds.

If the limiter output, which is also the discriminator input, is represented by the vector $V'\underline{/\phi}$ as in a.c. circuit analysis at the single frequency ω_0, D' is just the time derivative of ϕ, and the discriminator output voltage becomes

$$kV'(d\phi/dt).$$

We suppose that this is the output voltage whether V' and $(d\phi/dt)$ are fixed or changing rapidly.

Note that $V'(d\phi/dt)$, to which the output voltage of a discriminator is proportional, is just the component of the rate of change of the vector $V'\underline{/\phi}$, representing the input to the discriminator, in the direction perpendicular to this vector. (Cf. Fig. 1.)

THE CASE OF NO LIMITER

Before we consider the effect of both the limiter and the discriminator on the i-f f-m carrier and noise, we may consider the somewhat simpler effect of a discriminator alone. We shall suppose that the input to the discriminator is the sum of a narrow-band noise (because the i-f amplifier is a narrow-band-pass filter) of r.m.s. voltage σ and a carrier of constant deviation D and constant amplitude C. It makes no difference

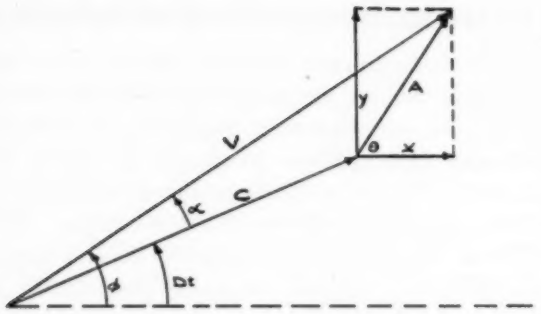


FIG. 1. Addition of carrier and noise vectors.

whether the deviation is always D or is merely D at the instant with which we are concerned, but it simplifies our nomenclature to assume it constant.

Thus, we may represent the carrier by the vector $C\angle Dt$, the noise by the vector $A\angle\theta = x+jy$, and their sum by the vector $V\angle\phi$. (See Fig. 1.) The angle between the carrier vector and the sum vector is designated as α. The discriminator output is $kV(d\phi/dt)$, which is k times the component of the rate of change of the sum vector in the direction perpendicular to the sum vector.

The rate of change of the carrier vector is CD and is directed perpendicular to the carrier vector. Its contribution to the discriminator output is, therefore, $kCD\cos\alpha$. The rate of change of the real component of the noise vector contributes $-k\dot{x}\sin\phi$, and that of the imaginary component $k\dot{y}\cos\phi$. Thus, the discriminator output voltage is

$$k(CD\cos\alpha - \dot{x}\sin\phi + \dot{y}\cos\phi). \qquad (1)$$

In order to find the mean and mean square values of (1), it is necessary to know a little about the statistics of the noise voltage.[10,14] These statistics depend on the spectrum of the noise. Since nearly all of the selectivity of the receiver lies in the i-f portion, and since most of the noise originates ahead of that portion, the noise at the output of the i-f amplifier has a spectral distribution of the same shape as the response *versus* frequency of the i-f amplifier, i.e., the amount of noise power in the frequency range between ω and $\omega+d\omega$ is proportional to the i-f power response (square of voltage response) at frequency ω. If the intermediate frequency, ω_0, is at the center (or, more generally, if the i-f response is not symmetric, at the centroid) of this i-f power response curve, the relevant statistics of the noise are quite simple and depend on only two quantities, the r.m.s. noise voltage σ and the radius of gyration of the i-f power-*versus*-frequency response curve ρ, defined by

$$\rho^2 = \int_0^\infty (\omega-\omega_0)^2 w(\omega)d\omega \Big/ \int_0^\infty w(\omega)d\omega,$$

where $w(\omega)$ is the power response of the i-f amplifier at frequency ω; ρ is a measure of the noise band width and so, also, of the rapidity of change of the vector representing the noise. The $(\omega-\omega_0)^2$ comes from differentiation with respect to time.

The necessary statistics are then the following.[10,14] The four quantities x, y, $\dot{x}$, and $\dot{y}$ are statistically independent. The mean value of each is 0. The mean square value of each of the first two is σ^2, and of the latter two $\rho^2\sigma^2$, and all are distributed normally (Gaussianly). It follows that the probability that A lies between A and $A+dA$ and θ lies between θ and $\theta+d\theta$ is

$$(2\pi)^{-1}d\theta\times\sigma^{-2}A\exp-(A^2/2\sigma^2)dA. \qquad (1a)$$

Since x, y, $\dot{x}$, and $\dot{y}$ are independent, $\dot{x}$ and $\dot{y}$ are independent of ϕ and α. Thus, we may average separately with respect to $\dot{x}$ and $\dot{y}$. Since the average value of each is zero, the average discriminator output is seen to be

$$kCD\langle\cos\alpha\rangle_{Av}, \qquad (2)$$

where $\langle\cos\alpha\rangle_{Av}$ means the average value of $\cos\alpha$. Squaring (1) and averaging, we get for the mean square discriminator output

$$k^2(C^2D^2\langle\cos^2\alpha\rangle_{Av}+\rho^2\sigma^2), \qquad (3)$$

where $\langle\cos^2\alpha\rangle_{Av}$ means the average value of $\cos^2\alpha$.

The mean value of $\cos\alpha$ is independent of the deviation D of the carrier, for the angle θ of the noise vector is equally likely to have any value between 0 and 2π, and so the angle Dt of the carrier vector has no effect on the distribution of the values of the angle α, although it may have some effect on the rapidity of the fluctuation of α. Thus, to simplify the treatment, we may, for purposes of computing the mean and mean square values of $\cos\alpha$, take $D=0$. With $D=0$, Fig. 1 becomes Fig. 2, and $\phi=\alpha$.

We may convert (1a) into the joint probability

that V is V within dV and ϕ is ϕ within $d\phi$ by writing $(C^2+V^2-2CV\cos\phi)$ (law of cosines) for A^2, and substituting the element of area $Vd\phi dV$ for $Ad\theta dA$,

$$\frac{1}{2\pi\sigma^2}\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)] \times Vd\phi dV. \quad (4)$$

The mean value of $\cos\alpha$, which is now $\cos\phi$, is obtained by integrating the product of $\cos\phi$ times the probability over all possible values of V and ϕ;

$$\langle\cos\alpha\rangle_{Nn} = \int_0^\infty \oint \cos\phi \times \frac{\exp[-(C^2+V^2-2CV\cos-)/(2\sigma^2)]}{2\pi\sigma^2} Vd\phi dV$$

$$= \frac{1}{\sigma^2}\int_0^\infty V\exp[-(C^2+V^2)/(2\sigma^2)]I_1(CV/\sigma^2)dV,$$

by the integral representation of I_1, the Bessel function of the first kind of order 1 and "imaginary argument." By the use of Eq. (5), p. 394, of Watson[16] after integrating by parts and substituting for the derivative of I_1,

$$dI_1(u)/du = \tfrac{1}{2}[I_0(u)+I_2(u)],$$

we finally arrive at

$$\langle\cos\alpha\rangle_{Nn} = \frac{C}{2\sigma}\left(\frac{\pi}{2}\right)^{\frac{1}{2}}\exp\left(-\frac{C^2}{4\sigma^2}\right) \times\left[I_0\left(\frac{C^2}{4\sigma^2}\right)+I_1\left(\frac{C^2}{4\sigma^2}\right)\right]$$

$$= \frac{(\pi)^{\frac{1}{2}}}{2}X\exp\left(-\frac{X^2}{2}\right) \times\left[I_0\left(\frac{X^2}{2}\right)+I_1\left(\frac{X^2}{2}\right)\right], \quad (5)$$

where

$$X = \frac{C}{\sqrt{2}\sigma}$$

is the input r.m.s. carrier-to-noise ratio. The functions $e^{-u}I_0(u)$ and $e^{-u}I_1(u)$ are tabulated in the back of Watson.

[16] G. N. Watson, *Bessel Functions* (The Macmillan Company, New York, 1944), revised edition.

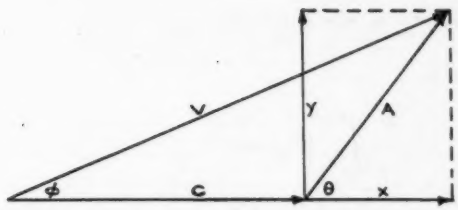


FIG. 2. Addition of carrier and noise vectors when $D=0$.

Thus we have the d.c. component of the discriminator output $kCD\langle\cos\alpha\rangle_{Nn}$, when the carrier deviation is D; expression (2) is evaluated. If the carrier is modulated sinusoidally at frequency ω_m to a maximum deviation D_0, with $D=D_0\cos\omega_m t$, the discriminator output contains the component $kCD\langle\cos\alpha\rangle_{Nn}\cos\omega_m t$, i.e., an output signal component of r.m.s. value

$$\frac{1}{\sqrt{2}}kCD_0\langle\cos\alpha\rangle_{Nn} \quad (6)$$

and no distortion, inasmuch as the d.c. discriminator output is strictly proportional to D.

We turn now to the problem of finding the mean square noise output. When D varies sinusoidally as above, we must average this mean square discriminator output (3) over the variation in D, which means merely that we replace D^2 by its mean value, $(1/2)D_0^2$. The mean square discriminator output then becomes

$$k^2[(1/2)C^2D_0^2\langle\cos^2\alpha\rangle_{Nn}+\rho^2\sigma^2], \quad (7)$$

of which it remains to evaluate $\langle\cos^2\alpha\rangle_{Nn}$, which we do in the same way as $\langle\cos\alpha\rangle_{Nn}$.

$$\langle\cos^2\alpha\rangle_{Nn} = \frac{1}{2\pi\sigma^2}\int_0^\infty \oint \cos^2\phi \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]Vd\phi dV$$

$$= \frac{1}{2\pi\sigma^2}\int_0^\infty \oint (1-\sin\phi\sin\phi) \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]Vd\phi dV.$$

Because (4) is a normalized probability, the integral of the first term on the right is unity. Integrating the second term by parts with

respect to ϕ, we thus get

$$\langle\cos^2\alpha\rangle_{Av}$$

$$=1-\frac{1}{2\pi C}\int_0^\infty \oint \cos\phi \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]d\phi dV$$

$$=1-\frac{\sigma^2}{2\pi C^2}\oint\int_0^\infty\left(\frac{C\cos\phi-V}{\sigma^2}+\frac{V}{\sigma^2}\right) \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]dVd\phi$$

$$=1-\frac{\sigma^2}{C^2}\exp\left(-\frac{C^2}{2\sigma^2}\right)+\frac{\sigma^2}{C^2},$$

in integrating the last term of which we have again utilized the normality of (4). Substituting X for $C/\sqrt{2}\sigma$, we have, finally,

$$\langle\cos^2\alpha\rangle_{Av}=1-[1-\exp(-X^2)]/(2X^2). \quad (8)$$

Substituting (8) into (7), we obtain the explicit expression for the mean square discriminator output. To obtain the mean square *noise* output we must subtract the mean square signal output, $(1/2)k^2C^2D_0^2\langle\cos\alpha\rangle_{Av}^2$; the result is

$$(1/2)k^2C^2D_0^2(\langle\cos^2\alpha\rangle_{Av}-\langle\cos\alpha\rangle_{Av}^2)+k^2\rho^2\sigma^2.$$

With $D_0=\rho$, which is about the largest value D_0 can have if the f-m carrier is not to be distorted by the cutting off of important sidebands by the i-f amplifier, our result for the noise becomes

$$k^2\rho^2\sigma^2[1+X^2(\langle\cos^2\alpha\rangle_{Av}-\langle\cos\alpha\rangle_{Av}^2)]. \quad (9)$$

It is evident that the noise output increases with modulation of the carrier. When the carrier is either very weak or very strong, this increase is small, in the former case because X is small, and in the latter because $\cos\alpha$ is always so close in value to 1 that the difference between its mean square value and the square of its mean value is unimportant.

SIGNAL OUTPUT WITH A LIMITER

The expressions obtained above for the discriminator signal and noise outputs are readily extended to the case where there is a limiter ahead of the discriminator. During the instants when the amplitude of the input to the limiter V is less than V_0, i.e., when the limiter is not limiting, the instantaneous discriminator output for fixed carrier deviation D is

$$k(V_1/V_0)(CD\cos\alpha-\dot{x}\sin\phi+\dot{y}\cos\phi),$$

which is just (V_1/V_0), the voltage gain of the limiter, times what we had before without a limiter. For fixed α its average value is

$$kCD(V_1/V_0)\cos\alpha.$$

During the instants when $V>V_0$, i.e., when the limiter is limiting, its effect is to introduce a smaller voltage gain (V_1/V). Thus, under these circumstances, for given V and α, the mean discriminator output is

$$kCD(V_1/V)\cos\alpha.$$

Hence, the mean discriminator output voltage when V and α are allowed to take on all possible values is kCD times the mean value of a quantity which is $(V_1/V_0)\cos\alpha$ when $V<V_0$ and is $(V_1/V)\cos\alpha$ when $V>V_0$. To calculate this mean we may again assume that $D=0$, $\alpha=\phi$. Again using (4) and the integral definition of I_1, we have for the mean discriminator output voltage

$$kCD\frac{1}{2\pi\sigma^2}\left[\int_0^{V_0}\oint\frac{V_1}{V_0}V\cos\phi \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]d\phi dV + \int_{V_0}^\infty\oint V_1\cos\phi \times\exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]d\phi dV\right]$$

$$=kCD\frac{1}{\sigma^2}\left[\int_0^{V_0}\frac{V_1}{V_0}V\exp[-(C^2+V^2)/(2\sigma^2)] \times I_1\left(\frac{CV}{\sigma^2}\right)dV+\int_{V_0}^\infty V_1 \times\exp[-(C^2+V^2)/(2\sigma^2)]I_1\left(\frac{CV}{\sigma^2}\right)dV\right]$$

$$=kCD\frac{V_0V_1}{\sigma^2}\left[\int_0^1 u\exp\left(-\frac{C^2}{2\sigma^2}-\frac{V_0^2}{2\sigma^2}u^2\right) \times I_1\left(\frac{CV_0}{\sigma^2}u\right)du+\int_1^\infty \exp\left(-\frac{C^2}{2\sigma^2}-\frac{V_0^2}{2\sigma^2}u^2\right)I_1\left(\frac{CV_0}{\sigma^2}u\right)du\right]$$

$$= 2\sqrt{2}k\sigma D\frac{V_1}{V_0}Xz^2\exp(-X^2)$$

$$\times\left[\int_0^1 u\exp(-z^2u^2)I_1(2Xzu)du\right.$$

$$\left.+\int_1^\infty \exp(-z^2u^2)I_1(2Xzu)du\right],$$

where $z = V_0/\sqrt{2}\sigma$.

Hence, when the carrier is modulated sinusoidally with amplitude D_0, the discriminator output contains a signal component of r.m.s. value

$$2k\sigma D_0\frac{V_1}{V_0}Xz^2\exp(-X^2)$$

$$\times\left[\int_0^1 u\exp(-z^2u^2)I_1(2Xzu)du\right.$$

$$\left.+\int_1^\infty \exp(-z^2u^2)I_1(2Xzu)du\right] \quad (10)$$

and no distortion. Unfortunately, these integrals cannot be evaluated in terms of tabulated functions; it is necessary to resort to numerical integration. The second integral was converted into an integral from 0 to 1 by substitution for $1/u$, and Gauss's seven-point quadrature formula was applied to both integrals in the two cases $z=1$, $X=1$, 2 (carrier 0 and 6 db above the noise).

The case of no limiter is represented by infinite z and $V_1 = V_0$. As z becomes infinite, therefore, (10) approaches (V_1/V_0) times (6), with (5) substituted into it. Thus, (10) can be plotted for infinite z.

Figure 3 is a plot of (10); it was plotted on the basis of the curve for $z=\infty$, the points $z=1$, $X=1$, 2, and the asymptotes to all of the curves.

When $2Xzu$ is small compared to unity, $I_1(2Xzu)$ is approximately Xzu; when it is large, $I_1(2Xzu)$ is approximately $\exp(2Xzu)/2(\pi Xzu)^{\frac{1}{2}}$. Inserting the former approximation in both integrands of (10), we obtain as the r.m.s. discriminator signal output, when the input carrier-to-noise ratio is small ($X \ll 1$),

$$k\sigma D_0\frac{V_1}{V_0}X^2\int_0^z \exp(-u^2)du.$$

This integral, which is one variety of the error function Erf(z), has been tabulated often and variously. This expression gives us the left asymptotes to the curves in Fig. 3 corresponding to various degrees of limiting (z's).

The asymptotes on the right are obtained by using the other approximation. When $X \gg z+1$, the limiting is nearly complete, and the first integral in (10) is negligible compared to the second, whose integrand is now a sharply peaked (Gaussian) function, being approximately proportional to $\exp-(zu-X)^2$. Thus, the integration is reduced essentially to one of

$$\int_{-\infty}^{\infty}\exp(-t^2)dt = (\pi)^{\frac{1}{2}},$$

and we have as the r.m.s. discriminator signal output component, when $X \gg z+1$,

$$\frac{V_1}{V_0}k\sigma D_0 z.$$

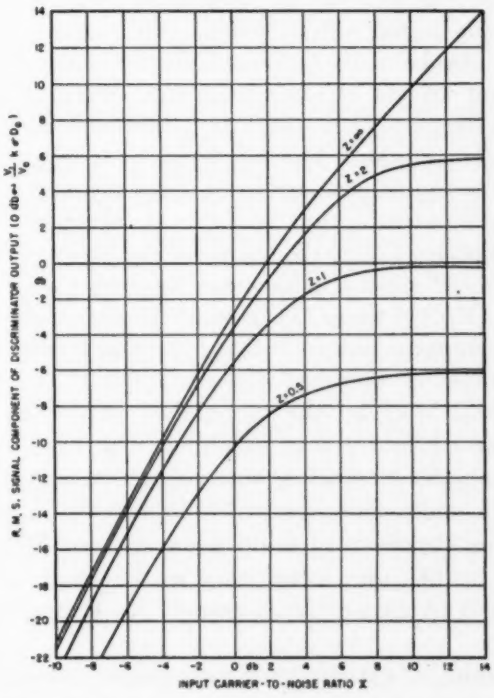


FIG. 3. Discriminator signal output *vs.* carrier strength for several values of z.

If $z \gg X \gg 1$, the peak of the product occurs within the range of the first integral of (10), so that the *second* integral becomes negligible by comparison. Here we find for the r.m.s. discriminator signal output component

$$\frac{V_1}{V_0} k\sigma D_0 X.$$

This gives the asymptote to the curve for infinite z.

We may also consider the behavior of (10) when $z \ll 1$, i.e., with extreme limiting. The first integral is then negligible by comparison to the second, whose lower limit we may make 0 without altering its value significantly. The resulting integral can be integrated by means of Eq. (5), p. 394, of Watson.[16] Thus, it turns out that the r.m.s. discriminator signal output voltage is

$$(V_1/V_0) k\sigma D_0 z[1-\exp(-X^2)],$$

in exact agreement with Rice[14] and with Middleton,[15] Eq. (2.37).

NOISE OUTPUT WITH A LIMITER

The changes in the calculation of the no-limiting signal output which were made above may also be made in the calculation of the noise output to get the mean square discriminator noise output when there is a limiter ahead of the discriminator. During the instants when the limiter is not limiting ($V < V_0$) the mean square discriminator output voltage is obtained by multiplying (7) by the square of the limiter gain, $(V_1/V_0)^2$; when the limiter is limiting ($V > V_0$), it is $(V_1/V)^2$ times (7). Hence, the mean square discriminator output voltage, averaged over *all* instants, is obtained by multiplying (7) by $(V_1/V_0)^2$ and by the probability that V is V within dV and ϕ is ϕ within $d\phi$, which is given by (4), and integrating with respect to ϕ from 0 to 2π, and with respect to V from 0 to V_0, and adding the result of multiplying (7) by $(V_1/V)^2$ and by (4), and integrating with respect to ϕ from 0 to 2π and with respect to V from V_0 to ∞.

Taking $D=0$ again for purposes of computation of the "mean value" of $\cos^2\alpha$ of (7), which we may then write as $\cos^2\phi$, we have for the mean square discriminator output voltage

$$(1/2)k^2C^2D_0^2 \cdot \frac{1}{2\pi\sigma^2}\Big[\int_0^{V_0} \oint \left(\frac{V_1}{V_0}\right)^2 V \cos^2\phi \times \exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]\,d\phi\,dV + \int_{V_0}^{\infty} \oint \frac{V_1^2}{V} \cos^2\phi \times \exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]\,d\phi\,dV\Big]$$

$$+k^2\rho^2\sigma^2 \cdot \frac{1}{2\pi\sigma^2}\Big[\int_0^{V_0} \oint \left(\frac{V_1}{V_0}\right)^2 V \times \exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]\,d\phi\,dV + \int_{V_0}^{\infty} \oint \frac{V_1^2}{V} \times \exp[-(C^2+V^2-2CV\cos\phi)/(2\sigma^2)]\,d\phi\,dV\Big].$$

As in obtaining (8), we may write $\cos^2\phi$ as $(1-\sin\phi\sin\phi)$ and integrate partially the resulting second term. The first term is like the second bracket here with which we thus combine it to obtain, after performing the integration with respect to ϕ by means of the integral definitions of the Bessel functions,

$$\left(\frac{k^2C^2D_0^2}{2\sigma^2}+k^2\rho^2\right)\Big[\int_0^{V_0}\left(\frac{V_1}{V_0}\right)^2 V \times \exp[-(C^2+V^2)/(2\sigma^2)]I_0\left(\frac{CV}{\sigma^2}\right)dV + \int_{V_0}^{\infty} \frac{V_2^2}{V} \exp[-(C^2+V^2)/(2\sigma^2)]I_0\left(\frac{CV}{\sigma^2}\right)dV\Big]$$

$$-(1/2)k^2C^2D_0^2\Big[\int_0^{V_0}\left(\frac{V_1}{V_0}\right)^2 \times \exp[-(C^2+V^2)/(2\sigma^2)]I_1\left(\frac{CV}{\sigma^2}\right)dV + \int_{V_0}^{\infty} \frac{V_1^2}{V^2} \exp[-(C^2+V^2)/(2\sigma^2)]I_1\left(\frac{CV}{\sigma^2}\right)dV\Big].$$

Substituting X for $C/\sqrt{2}\sigma$ and z for $V_0/\sqrt{2}\sigma$, we obtain for the mean square discriminator output

voltage

$$2\left(\frac{V_1}{V_0}\right)^2\left(1+\frac{D_0^2}{2}X^2\right)k^2\rho^2\sigma^2z^2\exp(-X^2)$$
$$\times\left[\int_0^1 u\exp(-z^2u^2)I_0(2Xzu)du\right.$$
$$\left.+\int_1^\infty\frac{1}{u}\exp(-z^2u^2)I_0(2Xzu)du\right]$$
$$-\left(\frac{V_1}{V_0}\right)^2k^2\rho^2D_0^2Xz\exp(-X^2)$$
$$\times\left[\int_0^1\exp(-z^2u^2)I_1(2Xzu)du\right.$$
$$\left.+\int_1^\infty\frac{1}{u^2}\exp(-z^2u^2)I_1(2Xzu)du\right]. \quad (11)$$

The Case of $D_0=0$

To get the mean square *noise* output we must subtract the square of (10), which is the r.m.s. signal output, from (11). In case $D_0=0$ there is no signal output, and (11) simplifies considerably, too. Figure 4 is a plot of (11) for $D_0=0$. It was drawn on the basis of two points calculated by the aid of Gaussian seven-point integration, $z=1$, $X=1$, 2, having ordinates -1.75 db and -5.61 db, respectively, the curve (?) for large (infinite) z, and the asymptotes to the curves, which are investigated below.

When $z\gg X+1$, the limiter hardly ever limits, and, with $D_0=0$, the main contribution to (11) comes from small u's; the second integral is negligible. Since the integrand of the first integral has already become very small by the time u has reached 1, we may extend the first integration to ∞ without any appreciable difference. The integration can then be performed with the aid of Eq. (1), p. 393, of Watson.[16] Thus, the mean square discriminator output with $z\gg X+1$ is

$$\left(\frac{V_1}{V_0}\right)^2k^2\rho^2\sigma^2,$$

in agreement with (7) and (9). It is independent of the carrier amplitude because the carrier vector, when constant, has no effect on the rate of change of the sum of the carrier and noise vectors.

Let us determine the mean square discriminator output for $D_0=0$ with X small. Obviously, when X is small the carrier has little influence on the discriminator output, and the output noise is practically the same as it is with the carrier entirely absent. Thus, we set $X=0$ in (11) and obtain immediately

$$\left(\frac{V_1}{V_0}\right)^2k^2\rho^2\sigma^2[1-\exp(-z^2)-z^2Ei(-z^2)]. \quad (12)$$

The exponential integral Ei is tabulated at the beginning of Jahnke and Emde.[17]

When X is large and, in particular, when $X\gg z+1$, nearly all of the discriminator output is due to limiter input amplitudes which considerably exceed V_0. The principal contribution of the integrals to (11) comes from large u's. Thus, we may ignore the first integral and insert into the second the asymptotic expression for $I_0(2Xzu)$ for large argument which, like that for I_1, is $\exp(2Xzu)/2(\pi Xzu)^{\frac{1}{2}}$. Integrating as we did in the corresponding case in the preceding section, we obtain for the mean square discriminator output

$$\left(\frac{V_1}{V_0}\right)^2k^2\rho^2\sigma^2z^2/X^2. \quad (13)$$

We may also consider the limiting case of (11) in which $z\ll1$. Then again the first integral's contribution is negligible. If $X\ll1/z$, we may divide the second integral into two parts, in the first of which $2Xzu\ll1$, so that $I_0(2Xzu)$ and

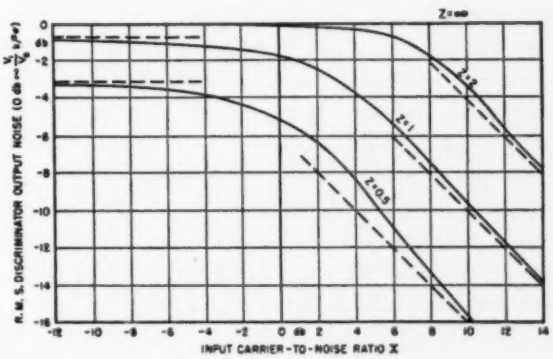


FIG. 4. Discriminator noise output *vs.* carrier strength for several values of z when carrier is unmodulated.

[17] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943).

$\exp(-z^2u^2)$ may be replaced by 1. Thus, if z_0 is such a number that $(1/X)\gg z_0\ll 1$, we have

$$2\left(\frac{V_1}{V_0}\right)^2 k^2\rho^2\sigma^2 z^2 \exp(-X^2)\left[\int_1^{z_0/z}\frac{du}{u} + \int_{z_0/z}^{\infty}\exp(-z^2u^2)I_0(2Xzu)\frac{du}{u}\right]$$

$$= 2\left(\frac{V_1}{V_0}\right)^2 k^2\rho^2\sigma^2 z^2 \exp(-X^2) \times\left[\log\frac{z_0}{z} + \int_{z_0}^{\infty}\exp(-u^2)I_0(2Xu)\frac{du}{u}\right].$$

The integral here does not depend on z, and so for any fixed X it is always possible to choose z sufficiently small that the integral is negligible by comparison to the logarithm, although this requires a z of the order of 10^{-10^5} for X of the order of unity. With this assumption, the mean square discriminator output becomes approximately

$$2(V_1/V_0)^2 k^2\rho^2\sigma^2 z^2 \exp(-X^2)\log(1/z), \quad (14)$$

since when z is very small $\log(1/z_0)\ll\log(1/z)$.

When $X\gg 1$, we may approximate the above integral, as usual, by replacing the Bessel function by the asymptotic approximation. Thus, we get for the mean square discriminator output, with the additional hypothesis that $X\gg 1$,

$$\left(\frac{V_1}{V_0}\right)^2 k^2\rho^2\sigma^2 z^2\left(2\exp(-X^2)\log\frac{1}{z} + \frac{1}{X^2}\right). \quad (15)$$

This quantity is seen to change rapidly from being approximately (14) to being approximately (13) as X increases through the value which makes its two terms equal, this value being about 4 for $z=10^{-10^5}$.

When $X\ll 1$ and $z\ll 1$, (14) and (12) should become equal. We see that they do by substituting the approximation for the exponential integral of small arguments into (12) and writing 1 for $\exp(-X^2)$ in (14).

The Case of $D_0=\rho$

When D_0 is not zero, both terms of (11) must be computed. Again, by use of Gauss's seven-point quadrature formula this has been done for the two cases $z=1$, $X=1$, 2. On account of the following relation, it is actually necessary to compute only three of the four definite integrals appearing in (11):

$$\int_0^1 \exp(-z^2u^2)I_1(2Xzu)du = \frac{\exp(-z^2)I_0(2Xz)-1}{2Xz} + \frac{z}{X}\int_0^1 u\exp(-z^2u^2)I_0(2Xzu)du.$$

It is easy to see from (11) or on physical grounds that, when the carrier is weak, the discriminator noise output is unaffected by modulation of the carrier. This is also true when the carrier is strong provided $z\gg X\gg 1$, as we saw before, and as one can verify by using *two* terms of the asymptotic expansions for the Bessel functions of large arguments in (10) and (11). The second terms are needed because the first terms give essentially only $\infty-\infty$. However, when $X\gg z+1$, such expansion shows that there is an increase in the mean square discriminator noise output in the ratio $(1+D_0^2/2\rho^2):1$.

We may also treat the case of $z\ll 1$ specially, in the same manner as we did with $D_0=0$. Provided $X\ll 1/z$, (11) becomes approximately

$$2\left(1+\frac{D_0^2}{2\rho^2}X^2\right)\left(\frac{V_1}{V_0}\right)^2 k^2\rho^2\sigma^2 z^2\exp(-X^2)\log\frac{1}{z},$$

or, with $D_0=\rho$,

$$2[1+(1/2)X^2]\left(\frac{V_1}{V_0}\right)^2 k^2\rho^2\sigma^2 z^2\exp(-X^2)\log\frac{1}{z}.$$

Thus, when $X\ll 1/z\gg 1$, the mean square discriminator output noise goes up in the ratio $[1+(1/2)X^2]:1$ when the carrier is modulated. It is unnecessary to subtract the mean square signal from (11) in this case because the former is very much smaller than the latter.

SIGNAL-TO-NOISE RATIO

The discriminator output signal-to-total-noise ratio is the quotient of (10) divided by the square root of the difference between (11) and the square of (10). It has been plotted with $D_0=\rho$ in Fig. 5. The curve for infinite z is obtained from (5)–(8).

A comparison of our signal and noise results

shows that for $X \ll 1$ the output signal-to-total-noise ratio is

$$\frac{\int_0^z \exp(-t^2)dt}{[1-\exp(-z^2)-z^2Ei(-z^2)]^{\frac{1}{2}}} X^2,$$

which becomes

$$\tfrac{1}{2}(\pi)^{\frac{1}{2}}X^2 \qquad (16)$$

for $z \gg 1$.

For $X \gg z+1$, the signal-to-total-noise ratio is

$$(\tfrac{2}{3})^{\frac{1}{2}}X; \qquad (17)$$

for $z \gg X \gg 1$, it is,

$$X.$$

When $z \ll 1$ and $X \ll 1/z$, the signal-to-total-noise ratio is

$$\frac{1-\exp(-X^2)}{\{2\exp(-X^2)[1+(1/2)X^2]\log 1/z\}^{\frac{1}{2}}} = \frac{\sinh \tfrac{1}{2}X^2}{(\tfrac{1}{2}[1+(1/2)X^2]\log 1/z)^{\frac{1}{2}}},$$

which is rapidly replaced by (17) as X increases through the value that makes the two terms of (15) equal.

For the sake of comparison, Fig. 5 includes the signal-to-total-noise curve for detection by a square law detector of a 100 percent amplitude-modulated carrier,[9,11,12]

$$X^2\left(\frac{2}{1+3X^2}\right)^{\frac{1}{2}},$$

which becomes (17) for large X and is 4.1 db better than (16) for small X.

In measuring the output signal-to-noise ratio of a receiver, one must bear in mind that the modulation of the carrier affects not only the signal output but also the noise output, and so, in order to measure the noise output, it is necessary to modulate the carrier sinusoidally and use a sharp filter which will pass only the corresponding signal output. The mean square noise output is then the difference between the mean square signal output, determined with the aid of the filter, and the mean square total output, measured without filtering.

It is to be emphasized that the foregoing

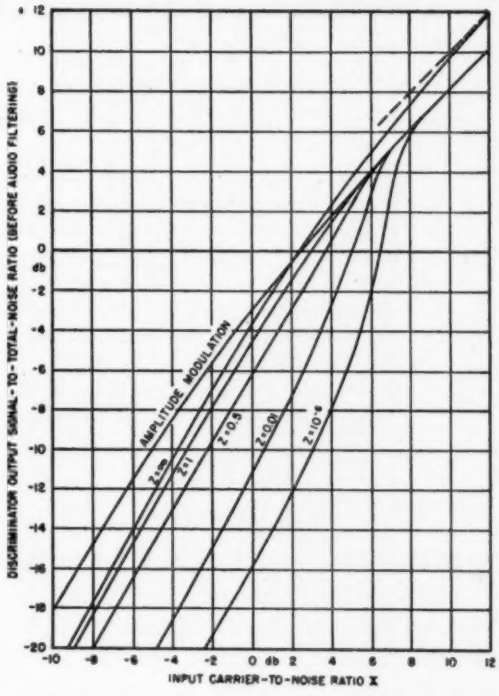


FIG. 5. Discriminator output signal-to-total-noise ratio *vs.* carrier strength for several values of z.

results pertain only to the unfiltered discriminator output noise. In the case of narrow-band f-m, the discriminator output noise has relatively few high frequency components, and so the subsequent filtering reduces only a little the amount of audible noise below the total output noise. However, in the case of wide-band f-m, most of the discriminator noise output is filtered out, and so it is quite necessary to know its spectral distribution. This has been done for the case of a large carrier-to-noise ratio and for the case of no limiting, in the latter case using a more realistic representation of the discriminator than here, in Cruft Laboratory Technical Report No. 31.* Middleton has obtained the general results.[15] The determination of the spectrum, of course, requires the use of much more complicated statistics than were required in the present paper.

It is a pleasure to acknowledge the inspiring association I have had with Dr. P. E. Le Corbeiller, the co-operation of Professor E. L. Chaffee, the guidance of Professor J. H. Van Vleck, and the interest of Dr. D. Middleton in this work.

Determination of the Elastic Constants of Solids by Ultrasonic Methods

WILLIAM C. SCHNEIDER AND CHARLES J. BURTON
Stamford Research Laboratories, American Cyanamid Company, Stamford, Connecticut
(Received June 21, 1948)

The application of ultrasonic methods to the determination of the elastic constants of solids is considered in some detail. It is shown that a rotating plate technique in which ultrasonic transmission is plotted as a function of the angle of incidence of the waves allows determination of the velocities of dilatation and shear waves in the plate. From these data, Poisson's ratio and the mechanical moduli may be determined. Details of an apparatus for making such measurements are given. The elastic constants of several metals have been measured with this equipment and the values obtained are shown to be in agreement with previous published data. In addition, measurements of a number of thermoplastic and thermosetting resins have been made successfully. It has been found that in the case of Melmac Resin 26-8B there is no variation in the elastic constants as a function of cure time.

INTRODUCTION

DURING the past few years, considerable interest has developed concerning methods for determining the elastic constants of plastics. These mechanical constants are of value for several reasons: (1) they provide a standard basis of comparison; (2) their changes upon the addition of plasticizers furnish a scientific criterion for studying plasticizing action; and (3) the effects of variations in production and forming techniques upon the ultimate mechanical properties can be followed.

Static methods of evaluating elastic constants have long been used in the study of materials such as the metals and glasses. Thermoplastic material in general exhibits appreciable plastic flow under mechanical stress and consequently considerable difficulty is encountered in interpretation of data obtained with the usual static techniques. Even in the case of thermosetting plastics, which are generally considered quite highly cured (i.e., extensively cross-linked), some care must be exercised when employing static techniques to minimize the effects of any slight plastic flow.

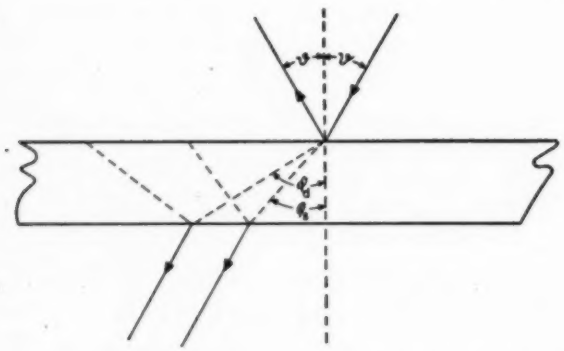


FIG. 1. Schematic representation of wave trains set up in a parallel plate. ϑ=angle of incidence=angle of reflection; φ_d=angle of refraction for dilatation waves; and φ_s=angle of refraction for shear waves.

At this point, it is desirable to discuss the nature of the elastic constants and to show their connection with vibration theory. In any isotropic solid, two types of deformation are possible upon application of a stress: (1) a simple elongation or contraction combined with a change in cross-sectional area, and (2) a relative sliding of planes parallel to the stress with no change in cross section. In the first case, the solid is said to be in a state of tension or compression and in the second case in a state of shear. Most deformations are a combination of the two, and, although tests have been developed which involve practically "pure" tension or shear, it is never possible to isolate the two completely.

The elastic constants or moduli express the resistance of a material to the above types of deformation. Young's modulus, E, expresses the resistance to extension and the modulus of rigidity, G, expresses the resistance to shear. In addition to these two constants, there is a third constant called Poisson's ratio, μ, which expresses the ratio of the cross-sectional contraction to the elongation in Type 1 deformations.

These three constants are related according to the following expression

$$E=2(1+\mu)G, \tag{1}$$

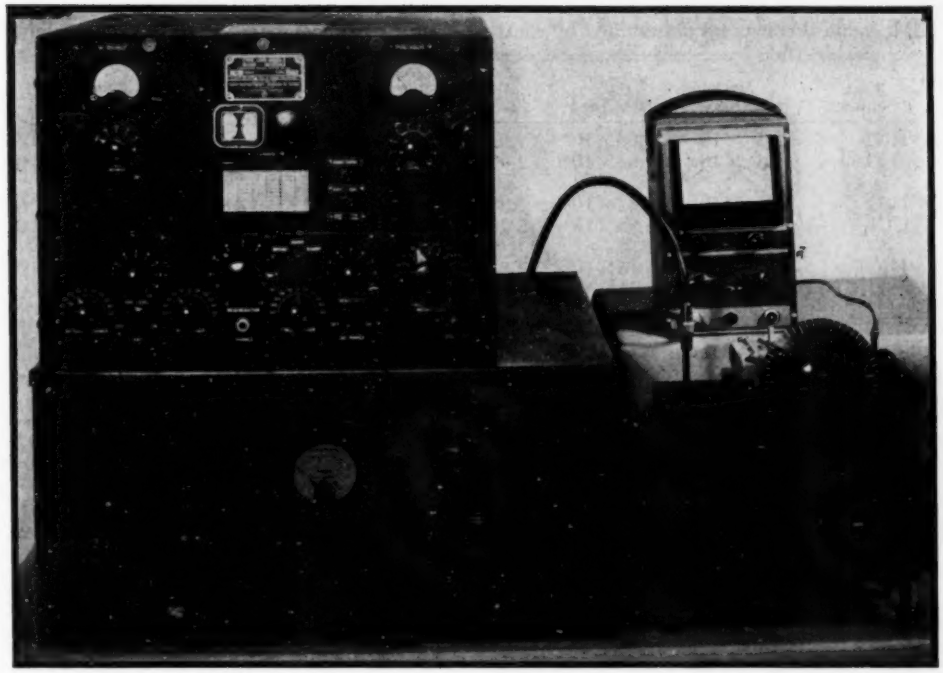


FIG. 2. Complete experimental arrangement. The oscillator and receiver are at the left and the measuring tank and output meter are at the right.

so that if any two of the constants are known, the third can be calculated.

When sound waves are propagated through a solid, they may travel with two different velocities depending upon which of the two types of deformation is associated with them. In the case of shear, the waves are known as shear, trans-

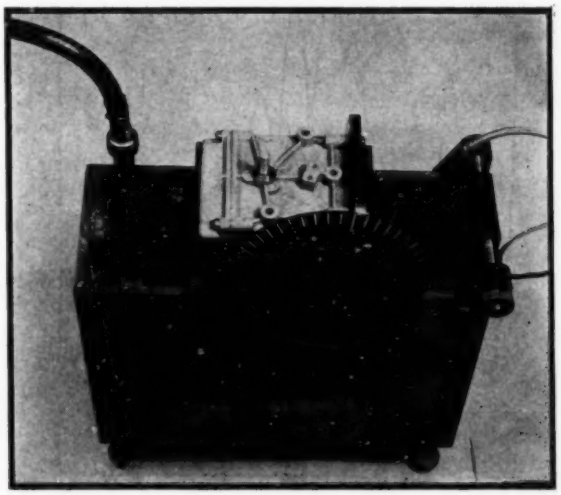
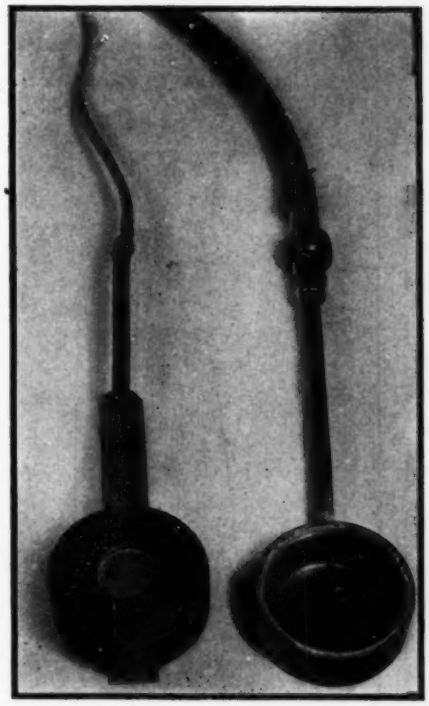


FIG. 3. Close-up of measuring tank with sample holder and crystals attached.



(a) (b)

FIG. 4. The receiving crystal: (a) front view; and (b) a rear view.

TABLE I. Mechanical constants determined by rotating plate method compared with previously published data.

Material	Tank liquid	V_l (cm/sec.)	V_d (cm/sec.)	V_s (cm/sec.)	$E\times10^{11}$ (dynes/cm²) obs.	handbook	$G\times10^{11}$ (dynes/cm²) obs.	handbook	μ	d in.
Aluminum	H_2O	1.5×10^5	7.05×10^5	2.82×10^5	6.45	6.96	2.14	2.37	0.40	~1/16
	H_2O	1.5×10^5	7.05×10^5	2.99×10^5	6.73	6.96	2.42	2.37	0.38	~1/8
Copper	H_2O	1.5×10^5	4.82×10^5	2.28×10^5	12.2	12.1–12.8	4.52	4.24	0.346	~1/8
	CCl_4	9.22×10^4	5.96×10^5	2.26×10^5	12.3	12.1–12.8	4.44	4.24	0.376	~1/8
Steel	H_2O	1.5×10^5	6.15×10^5	2.84×10^5	18.4	19.2	6.73	7.79	0.364	~1/16
	$C_6H_5CF_3$	9.85×10^4	6.31×10^5	2.72×10^5	17.1	19.2	6.16	7.79	0.386	~1/16

verse or ratational waves, and their velocity of propagation V_s is given by

$$V_s=(G/\rho)^{\frac{1}{2}}, \tag{2}$$

where ρ is the density of the material. When the deformation is a simple extension, the vibrations are known as longitudinal, dilatation, or irrotational waves and their velocity V_d in an infinite medium is given by

$$V_d=\left(\frac{(1-\mu)E}{\rho(1+\mu)(1-2\mu)}\right)^{\frac{1}{2}}. \tag{3}$$

TABLE II. Influence of thickness upon sound velocities and elastic constants for a series of aluminum sheets.

Sample	d (mm)	V_d (cm/sec.)	V_s (cm/sec.)	G (dynes/cm²)	E (dynes/cm²)	μ
A	0.813	6.53×10^5	2.55×10^5	1.75×10^{11}	4.94×10^{11}	0.410
B	1.04	6.34	2.62	1.85	5.18	0.397
C	1.33	7.29	2.80	2.12	6.02	0.413
D	2.11	6.73	2.95	2.35	6.49	0.382
E	3.26	6.73	2.95	2.35	6.49	0.382

Poisson's ratio is expressed as

$$\mu=[k^2-2/2(k^2-1)], \tag{4}$$

where $k=V_d/V_s$. Therefore, if one can determine

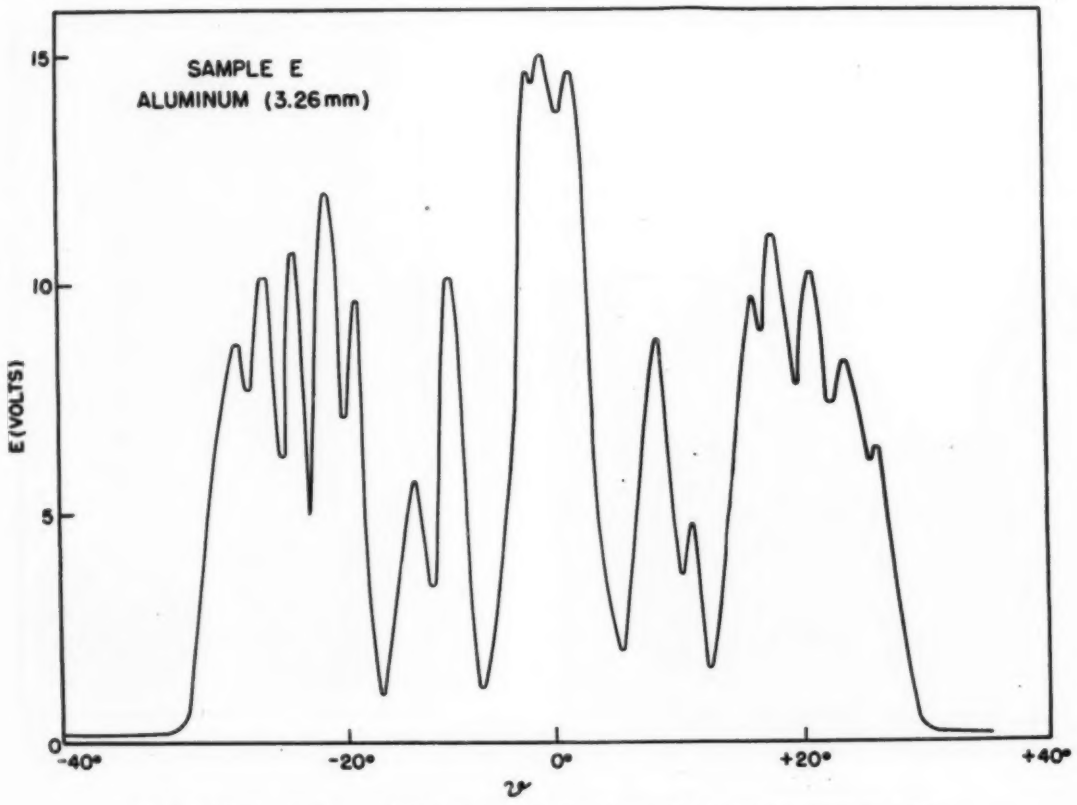


FIG. 5. Ultrasonic transmission pattern of aluminum (3.26 mm) in water at 5 Mc.

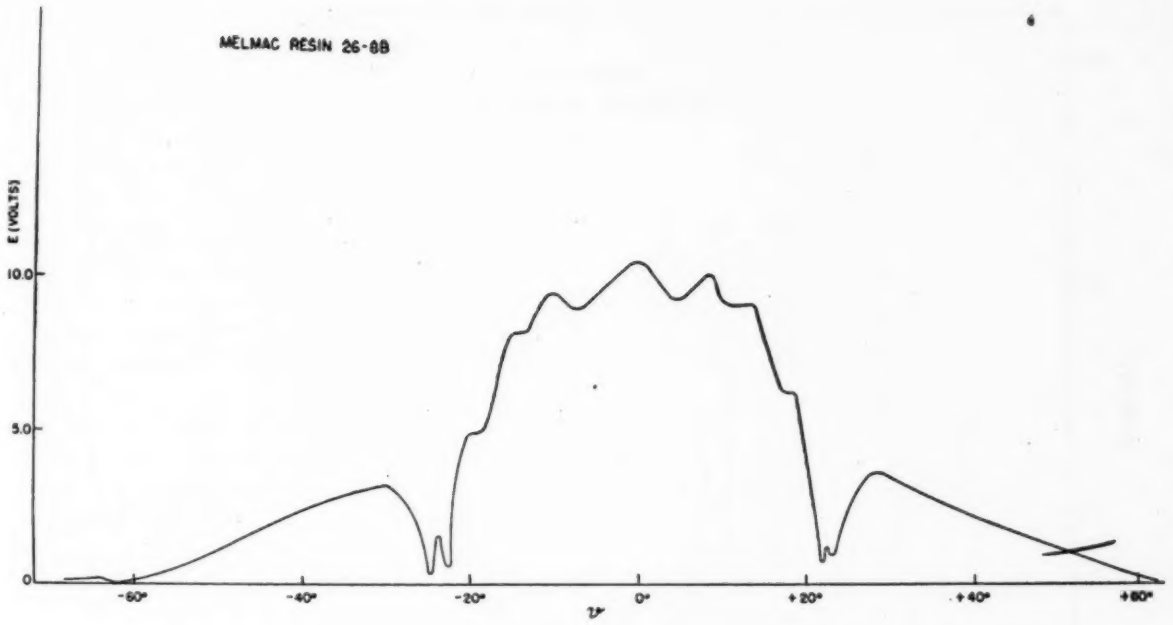


FIG. 6. Ultrasonic transmission pattern of Melmac resin 26–8B in water at 5 Mc.

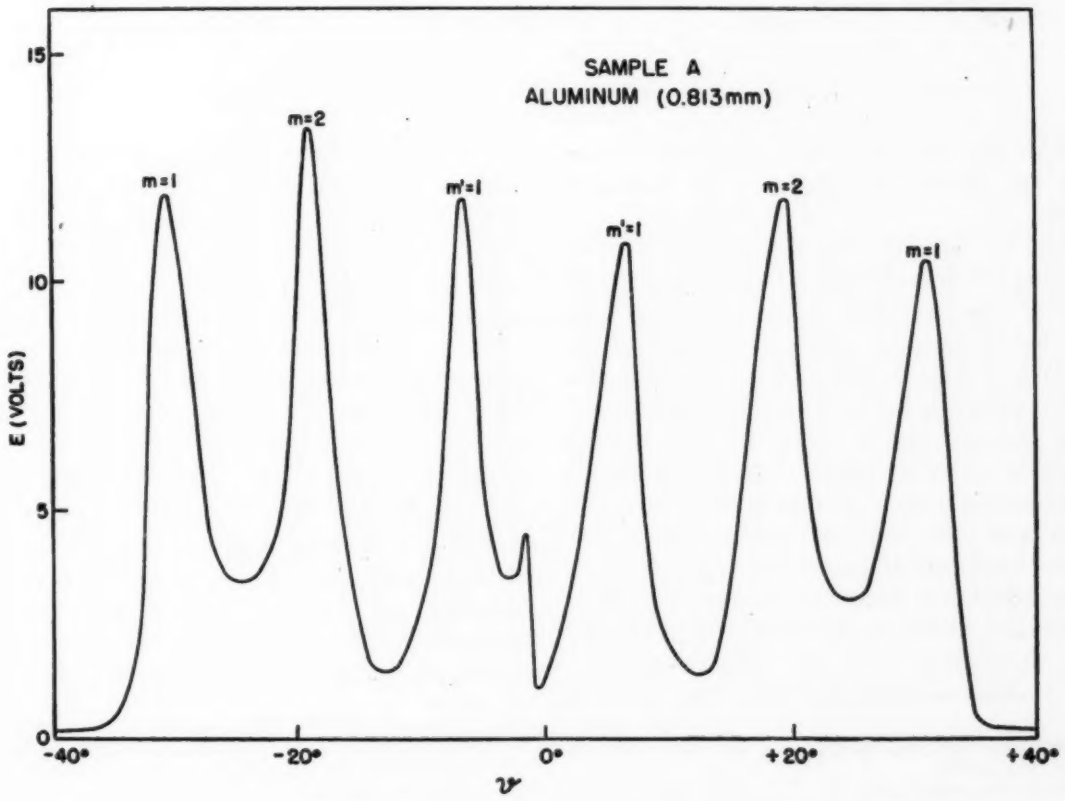


FIG. 7. Ultrasonic transmission pattern of aluminum (0.813 mm) in water at 5 Mc.

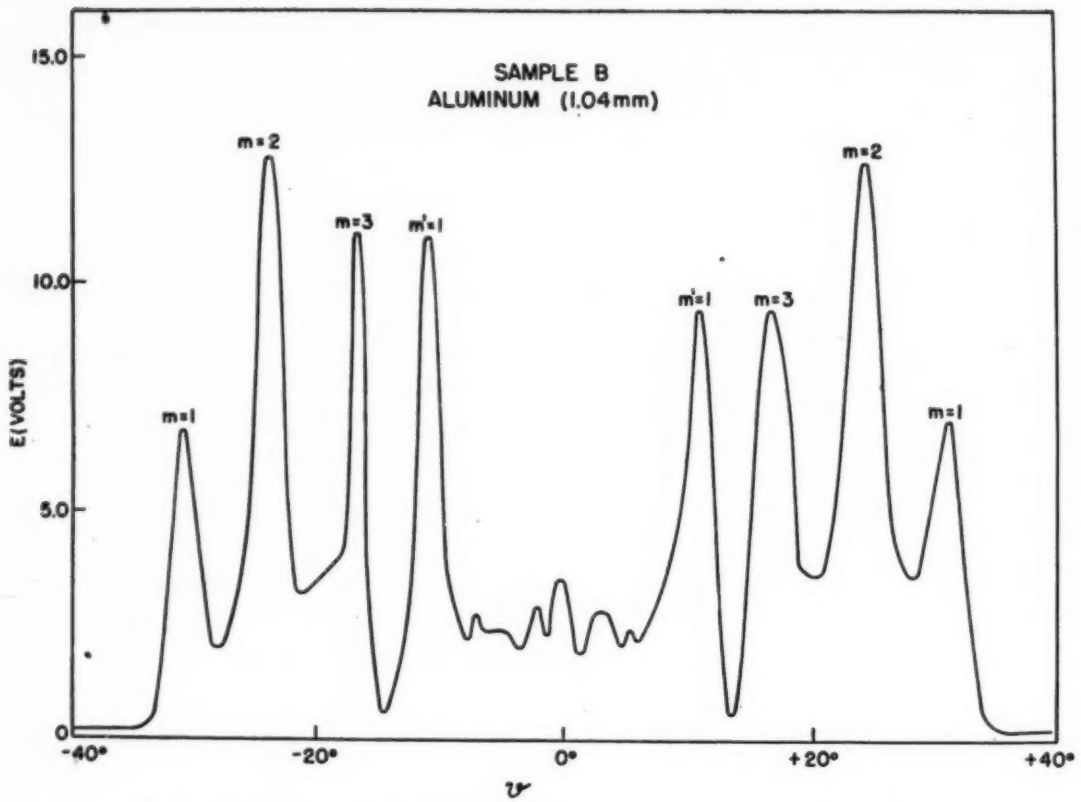


FIG. 8. Ultrasonic transmission pattern of aluminum (1.04 mm) in water at 5 Mc.

the velocities of the longitudinal and shear waves, the elastic constants may be evaluated at once.

There are several techniques available for measuring these velocities. One of particular interest is that suggested by Schaefer and Bergmann[1] for transparent solids which utilizes the diffraction of light by sound waves. A cube of a solid in vibration acts as a three-dimensional optical grating, and a beam of light passed through the vibrating solid will yield a characteristic diffraction pattern. For an isotropic material this pattern comprises two concentric circles, and the longitudinal and transverse velocities can be calculated from the measured radii of these circles. Although this method is relatively simple, it requires that the sample be transparent,[2] thus seriously limiting the field of application. Nevertheless, the method has been used with marked success in making systematic determinations of the elastic constants of optical glass.[3] More recently, the method has been applied in modified form to the study of transparent plastics by Szymanowski.[4]

Two other suitable techniques were first suggested by Bär and Walti[5] and by Bez-Bardili.[6] They are based upon variations in the transmission of sound waves through a plate. The waves are generated by a piezoelectric crystal and are formed into a beam by a suitable aperture. This beam impinges upon the sample plate, and the energy transmitted through the plate is detected in some suitable way, either by utilization of the Debye-Sears light diffraction technique as suggested by Bär and Walti[5] or by

[1] C. Schaefer and L. Bergmann, Sitz.-Ber. Berliner Akad., p. 155 (1934).

[2] Partially successful attempts have been made by C. Schaefer and L. Bergmann [Zeits. Tech. Phys. **17**, 441 (1936)] to modify the method for the investigation of opaque materials. The technique is similar to that already described except that the diffraction patterns are produced by reflection of light from a surface of the vibrating sample.

[3] C. Schaefer, L. Bergmann and H. J. Goehlich, Glastechn. Ber. **15**, 447 (1937).

[4] W. T. Szymanowski, J. App. Phys. **15**, 627 (1944).

[5] R. Bär and A. Walti, Helv. Phys. Acta. **7**, 658 (1934).

[6] W. Bez-Bardili, Zeits. Physik. **96**, 761 (1935).

a second piezoelectric crystal diametrically opposite the first.[7] In view of the extreme mismatch of acoustic impedance at a solid-gas interface, the generator, sample, and detector must be immersed in a liquid to obtain sufficient energy in the sound beam.

When the incident beam strikes the front surface of the sample plate, part is reflected and part is transmitted into the sample. Upon striking the rear surface of the plate, the beam is again partially reflected, the remainder of the energy being transmitted into the liquid and picked up by the detector. Accordingly, there are two wave trains existing in the sample: one transmitted through the front surface, and the other reflected from the rear surface. At normal incidence, these two wave trains give rise to a system of standing waves, and as the sample thickness is varied these trains alternately cancel and reinforce to give maximum and minimum transmission.

TABLE III. Interference orders and calculated velocities in a series of aluminum samples of increasing thickness.

Shear order	*A*	*B*	*D*
1	2.78×10^5	2.81×10^5	—
2	3.05	3.02	—
3	—	2.78	2.96×10^5
4	—	—	3.08
5	—	—	3.14
6	—	—	2.98
Longitudinal order			
1′	6.91×10^5	6.35×10^5	—
2′	—	—	6.39×10^5
3′	—	—	6.50×10^5

TABLE IV. Indices of refraction calculated from interference maxima compared with those obtained by direct measurement of angles of total reflection.

Copper	n_d	n_s
From angle of total reflection	0.309	0.652
From interference maxima	0.305	0.656

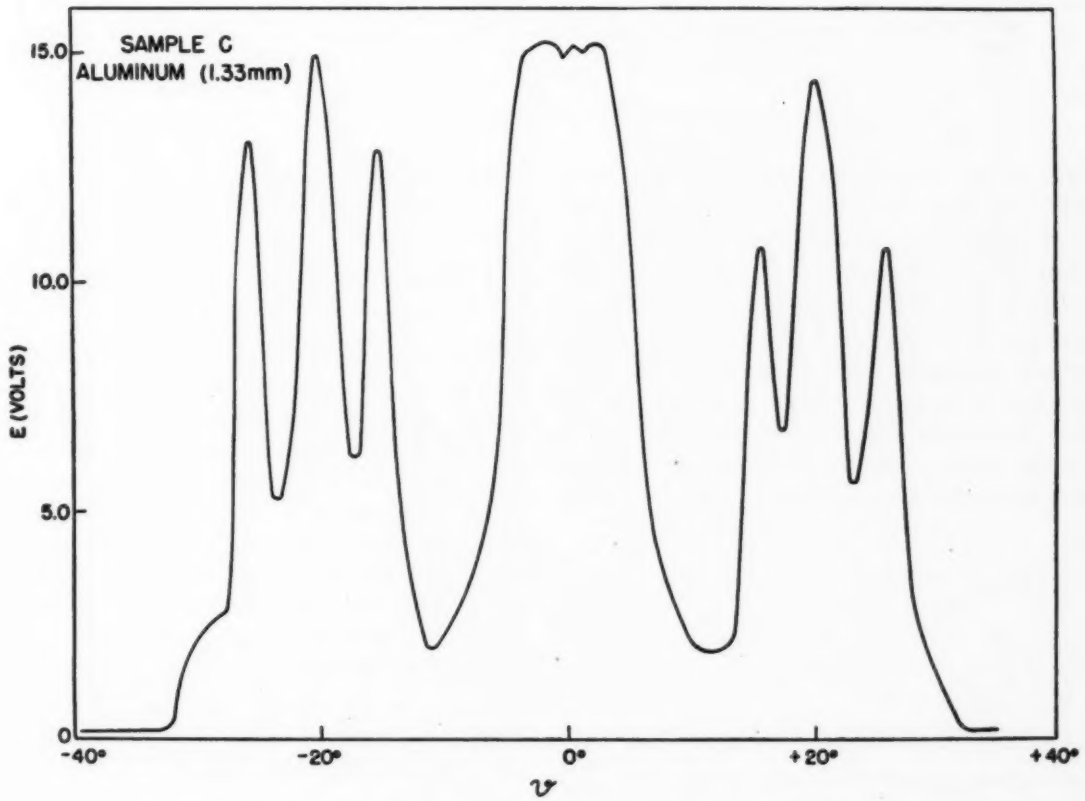


FIG. 9. Ultrasonic transmission pattern of aluminum (1.33 mm) in water at 5 Mc.

[7] D. Cochran and R. W. Samsel, G. E. Review, p. 39 (August, 1944).

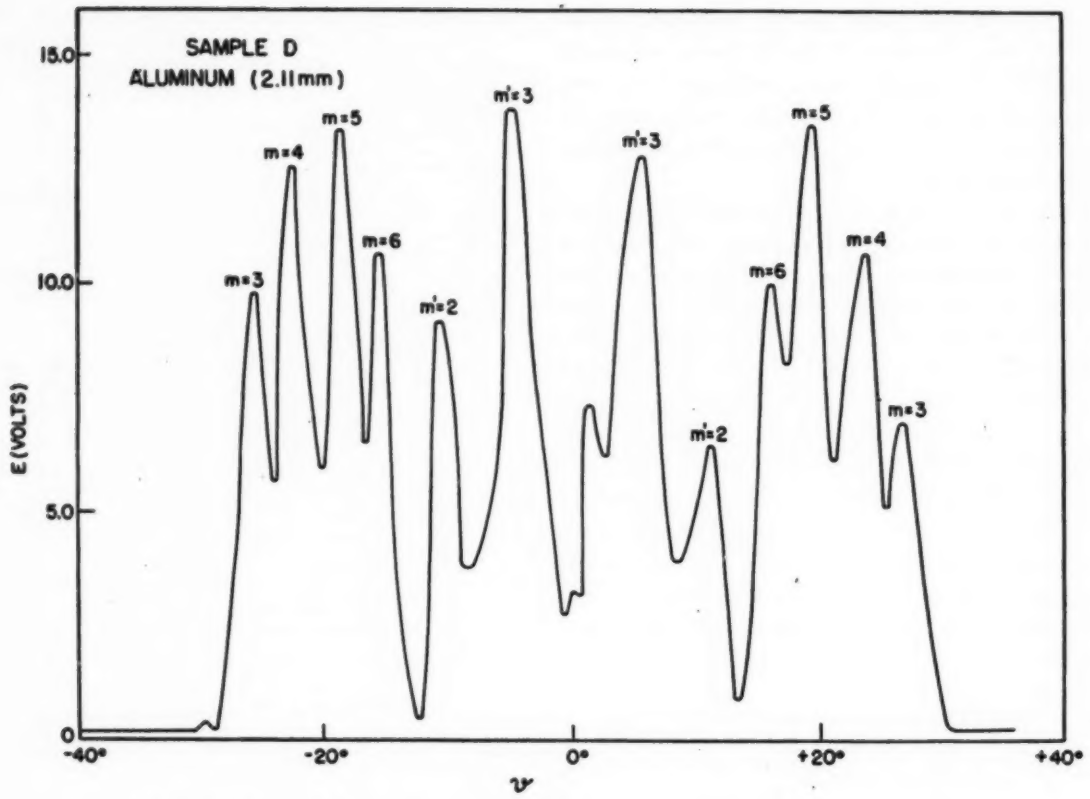


FIG. 10. Ultrasonic transmission pattern of aluminum (2.11 mm) in water at 5 Mc.

One of the techniques mentioned above depends upon this variation in intensity of transmitted energy with sample thickness. A wedge-shaped sample is used and is moved vertically through a horizontal sound beam. The distance of vertical travel required to give corresponding points of maximum (or minimum) transmission of energy is measured. Since the taper of the wedge is known, this vertical distance can be converted into a corresponding change in thickness which will be equal to one-half a wave-length of the incident vibration. Moreover, the frequency of the sound can be measured precisely, and the velocity of the waves can be calculated from the following relation

$$V=\lambda\nu, \tag{5}$$

where λ is the wave-length and ν the frequency.

The second technique mentioned above is based upon the variation in intensity of transmitted energy with the angle between the incident beam and the sample plate. A plane parallel sheet of material is used for the sample and is rotated about a vertical axis which is perpendicular to the horizontal sound beam. Figure 1 shows the wave trains excited when the beam strikes the sample plate at an angle ϑ. Since the velocity in the solid is usually greater than the velocity in the liquid, the wave trains in the solid are refracted away from the normal, and the following relations hold:[8]

$$\begin{aligned} n_d &= \sin\vartheta/\sin\varphi_d = V_l/V_d, \\ n_s &= \sin\vartheta/\sin\varphi_s = V_l/V_s, \end{aligned} \tag{6}$$

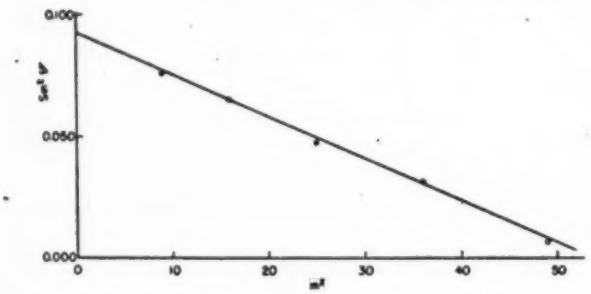


FIG. 11. Plot of Eq. (9) illustrating use of longitudinal interference maxima to determine index of refraction. Copper in water at 5 Mc. Sample thickness 3.2 mm.

[8] H. Reissner, Helv. Phys. Acta. 11, 140–155 (1938).

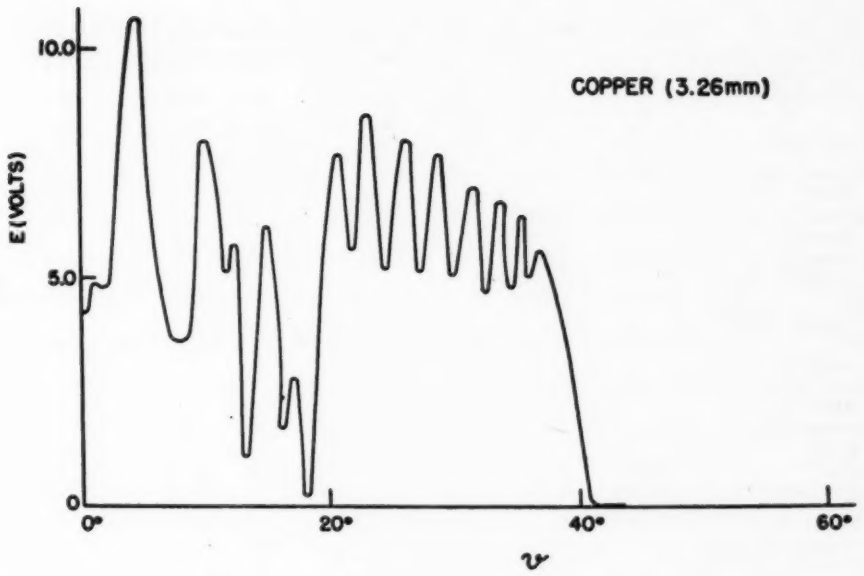


FIG. 12. Ultrasonic transmission pattern of copper (3.26 mm) in water at 5 Mc.

where n_d—index of refraction for dilatation waves, n_s—index of refraction for shear waves, V_l—velocity of sound in the liquid.

As ϑ increases, φ_d also increases until at a value ϑ_1, φ_d equals 90°. This means that the dilatation waves are totally reflected, and there is a pronounced minimum in the transmitted energy. As ϑ is increased further to a value ϑ_2, φ_s becomes 90°, and the shear waves are totally reflected. At ϑ_1 and ϑ_2 relations, (6) become

$$V_d=(V_l/\sin\vartheta_1), \quad V_s=(V_l/\sin\vartheta_2). \tag{7}$$

Accordingly, by determining experimentally the values of ϑ_1 and ϑ_2, and knowing the velocity of sound in the liquid, one can determine V_d and V_s by use of relations (7).

At intermediate values of ϑ, the transmitted energy goes through definite maxima and minima as a result of interference between incident waves and the wave trains within the sample. Unfortunately, these interference patterns can lead to considerable difficulty in locating the angle of total reflection for dilatation waves. In general, however, the intensity at the angle of total reflection is practically zero, and is much less than the interference minima.

It can be shown (5) that the positions of the interference maxima can be represented by the relation

$$m\lambda_d/2=d\cos\varphi_d \tag{8}$$

or

$$\sin^2\vartheta=n^2d-(\lambda_l/2d)^2m^2 \tag{9}$$

where λ_d—wave-length of dilatation waves, λ_l—wave-length in liquid, n_d—index of refraction for dilatation waves, m—order of the interference, d—sample thickness. Corresponding relations for the shear waves are

$$m\lambda_s/2=d\cos\varphi_s, \tag{10}$$

or

$$\sin^2\vartheta=n_s^2-(\lambda_l/2d)^2m^2. \tag{11}$$

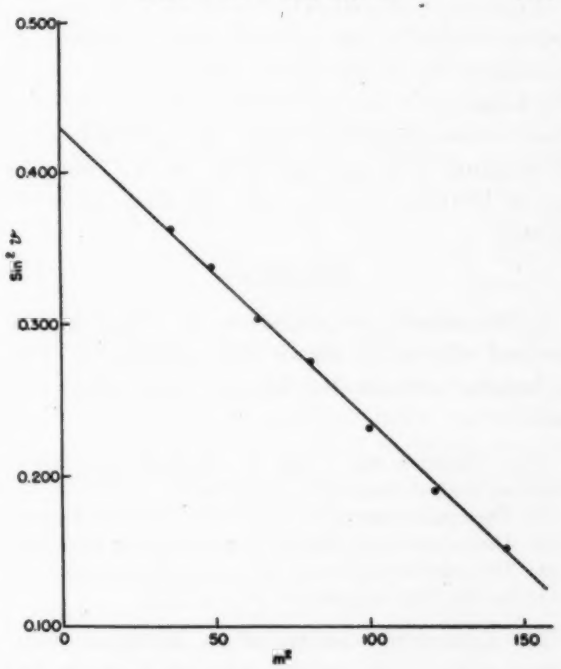


FIG. 13. Plot of Eq. (11) illustrating use of shear interference maxima to determine index of refraction. Copper in water at 5 Mc. Sample thickness 3.2 mm.

TABLE V. Elastic constants for various thermosetting* and thermoplastic** resins.

Material	Tank liquid	V_l $\times 10^5$ (cm/sec.)	V_d $\times 10^5$ (cm/sec.)	V_s $\times 10^5$ (cm/sec.)	F (dynes/cm²)	F (lbs./in.²)	G (dynes/cm²)	G (lbs./in.²)	μ	d (in.)
Melmac resin 1079*	$C_6H_5CF_3$	0.985	3.69	1.72	1.21×10^{11}	1.75×10^6	4.44×10^{10}	6.43×10^5	0.360	1/16
Melmac resin 26–8*B*	$C_6H_5CF_3$	0.985	3.69	1.72	1.21	1.75	4.44	6.43	0.360	1/8
Melmac resin *S*-6003*	$C_6H_5CF_3$	0.985	4.58	2.09	2.40	3.48	8.75	1.27×10^6	0.369	1/8
Glass (lantern slide)	$C_6H_5CF_3$	0.985	5.93	2.85	5.49	7.95	2.03×10^{11}	2.94	0.350	1/16
Glass (window)	$C_6H_5CF_3$	0.985	6.79	3.26	7.16	1.04×10^7	2.66	3.86	0.344	1/8
Beetle resin 20*RC*-4*	$C_6H_5CF_3$	0.985	3.96	1.74	1.26	1.83×10^6	4.55×10^{10}	6.60×10^5	0.381	1/8
Laminac resin 4116*	$C_6H_5CF_3$	0.985	2.52	1.29	4.83×10^{10}	7.00×10^5	1.83	2.67	0.322	1/8
Lucite**	CH_3OH	1.13	2.63	1.38	5.64	8.16	2.15	3.12	0.310	1/4
Polystyrene**	CH_3OH	1.13	2.30	1.18	3.91	5.66	1.48	2.14	0.322	1/4

* Thermosetting Resin.
** Thermoplastic Resin.

In addition to these types of interference, the dilatation and shear waves can interact with each other to give another series of maxima and minima whose positions can be represented by

$$m/d=\frac{\cos\varphi_d}{\lambda_d}+\frac{\cos\varphi_s}{\lambda_s}. \tag{12}$$

Equations (9) and (11) can be used to calculate sound velocities. If $\sin^2\vartheta$ is plotted against n^2, a straight line is obtained whose intercept gives the index of refraction of either the dilatation or shear waves depending upon the particular series of maxima and minima being used. Then, by use of relations (6) the velocities may be determined.

APPARATUS

In the present investigation, the rotating plate method discussed above was employed. This technique was selected for several definite reasons:

(1) A simple sample shape is employed so that no extensive sample preparation is required.
(2) The method does not involve the sample thickness.
(3) The sample is not damaged by the testing procedure.
(4) The experimental procedure is simple and rapid.
(5) No involved calculations are required.

The apparatus consists of an electronic oscillator used to drive the piezoelectric crystal producing the sound energy, a measuring tank, and a sensitive radio receiver used to amplify the energy picked up by the detector. Figure 2 is a photograph of the complete set-up. The oscillator and receiver are at the left and the measuring tank and output meter are at the right. Figure 3 is a closeup of the measuring tank with the sample holder and driving and receiving crystals attached. A National Type PW–O vernier radio dial is used to rotate the sample and allows a precision in angle measurement of about 0.4°. Figures 4(a) and 4(b) show front and back views respectively of the receiving crystal. The transmitting crystal is identical except that the front aperture is $\frac{1}{4}''$ in diameter. The electrodes were applied to the crystals by evaporating on a thin copper film, then electroplating copper to increase the electrode thickness. The crystal was attached to the holder with a low melting (98°C) solder, the front electrode being grounded to eliminate stray electrical pick-up.

EXPERIMENTAL

Although electrical apparatus was available for measurements at either 500 kc or 5 Mc, the present work was restricted to 5 Mc because the measuring tank was not suitable for low frequency work because of pronounced internal reflections of the sound energy.

In order to check the accuracy of the method, the elastic constants of several metals were measured and compared with published data.

Several different liquids were used as the tank liquid. Although the tank liquid does not affect the final results, it does influence considerably the angular position of the total reflection minima in the transmission curve. The lower the velocity of sound in the liquid, the smaller the angle of incidence for total reflection. Table I compares for several metals "handbook" elastic constants with those obtained in this investigation. As indicated by data of Willard,[9] the variation in V_{H_2O} with temperature is only ± 0.7 percent for a 5°C variation and can be neglected in comparison with the error in the critical angle. Consequently, the probable accuracy of these data is determined solely by the precision with which the critical angles can be evaluated. In the present case these angles can be determined to $\pm 0.072°$, and the corresponding deviation in $\sin\theta$, used in the calculations, is given by $d(\sin\theta) = \pm 0.072 \cos\theta$. Obviously, $d(\sin\theta)$ is a function of the angle and, within the range of values encountered can vary from ± 4.2 to ± 41 percent.

To illustrate the effects of the above deviation upon the final results, a sample calculation will be made for the copper sample in Table I. The critical angles for copper are $\vartheta_d = 18 \pm 0.072°$ and $\vartheta_s = 41 \pm 0.072°$ giving $d(\sin\vartheta_d) = \pm 0.0685$ and $d(\sin\vartheta_s) = \pm 0.0545$. Inserting these values in Eq. (7) and taking $V_{H_2O} = 1.498 \pm 0.005 \times 10^5$ cm/sec.[9] one obtains $V_d = 4.82 \pm 0.064 \times 10^5$ and $V_s = 2.28 \pm 0.050 \times 10^5$. Assuming the density of copper to be 8.70 ± 0.01, Eq. (2) yields, $G = 4.52 \pm 0.11 \times 10^{11}$ dynes/cm². Since the errors in the velocities effectively cancel when determining Poisson's ratio, Eq. (1) gives $E = 12.2 \pm 0.11 \times 10^{11}$ dynes/cm². The agreement is quite satisfactory. However, there seems to be a shift in the value of the shear modulus with sample thickness. Sample thickness does not enter explicitly into the calculations, but relations (7) are derived for an infinitely thick sample and are not strictly valid for extremely thin material. Although there is no simple expression relating sample thickness and accuracy of Eqs. (7), both Bez-Bardili[6] and Sanders[10] have shown empirically that if the product of sample thickness in centimeters and measuring frequency in megacycles is greater than 4 Mc-cm, relations (7) are valid.

TABLE VI. Elastic constants of various thermosetting resins as a function of cure time. *Sample 26–8B:* cellulose filled melamine-formaldehyde resin. *Sample 26–7A:* unfilled melamine-formaldehyde resin. *Sample 404–SE1102:* unfilled melamine-formaldehyde resin.

Material	Tank liquid	V_l $\times 10^5$ cm/sec.	V_d $\times 10^5$ cm/sec.	V_s $\times 10^5$ cm/sec.	E $\times 10^{11}$ dynes/cm²	G $\times 10^{11}$ dynes/cm²	μ	d in.
*26–8B**								
5′ 150°	CCl_4	0.922	3.59	1.69	1.16	0.428	0.358	1/8
10′ 150°	CCl_4	0.922	3.54	1.66	1.12	0.413	0.360	1/8
15′ 150°	CCl_4	0.922	3.54	1.66	1.12	0.413	0.360	1/8
2½′ 155°	CCl_4	0.922	3.63	1.66	1.13	0.413	0.367	1/16
7½′ 155°	CCl_4	0.922	3.59	1.67	1.14	0.419	0.362	1/16
*26–7A***								
5′ 155°	H_2O	1.49	3.33	1.62	1.06	0.394	0.348	1/16
10′ 155°	H_2O	1.49	3.23	1.61	1.02	0.389	0.333	1/16
*404–SE1102****								
3′ 155°	CH_3OH	1.13	3.34	1.61	0.945	0.350	0.349	1/16
6′ 155°	CH_3OH	1.13	3.34	1.61	0.945	0.350	0.349	1/16

* Cellulose filled melamine-formaldehyde resin.
** Unfilled melamine-formaldehyde resin.
*** Unfilled melamine-formaldehyde resin.

In Table II the observed velocities and derived elastic constants are tabulated for a series of aluminum samples of increasing thickness. It will be noted that variations in the elastic constants as a result of sample thickness disappear for thicknesses of 3 mm, or greater.

Figure 5 gives a typical experimental curve for a metal, and Fig. 6 shows the results for a typical plastic. The chief difference between the two is the number of maxima and minima in the curve for the metal. This results from the fact that there is less absorption (smaller acoustic loss) in the metal.

It was stated previously that the tank liquid influences considerably the positions of the reflection minima. This fact is useful because it allows a shifting of the experimental curve. For certain combinations of sample and tank liquid, the velocity of shear waves in the sample can become less than the velocity in the liquid. This causes the second of Eqs. (7) to become meaningless since the sine of the angle of incidence can never become greater than one. In other words, there is no angle of incidence for which total reflection occurs. An example is polystyrene in water. However, if methanol, which exhibits lower acoustic velocity than water, is used for the tank liquid, the angle of total reflection is shifted toward a smaller angle of incidence, and it is possible to obtain reflection. This is the

[9] G. W. Willard, J. Acous. Soc. Am. **19**, 237 (1947).
[10] F. H. Sanders, Canadian J. Rev. **17A**, 179 (1939).

chief value of changing the tank liquid. Unfortunately, the lowest reported acoustic velocity for liquids is about 9.22×10^4 cm/sec. for carbon tetrachloride. The shear velocity in lead and in many plastics is less than this value, and consequently these materials cannot be measured using the present technique. Air cannot be used as a tank liquid because there is almost 100 percent total reflection of sound energy at a solid-gas interface.

As stated above, the transmission curve for a metal exhibits two series of maxima and minima corresponding to interference of dilatation and shear waves, respectively. The number and position of these maxima vary with sample thickness. Although this effect can be explained theoretically, the situation can be illustrated in a simple manner by reference to Figs. 7–10 in which the transmission curves of a series of aluminum samples of increasing thickness are recorded. The small numbers adjacent to the interference maxima refer to the order of interference. In practice, the assignment of order numbers is a difficult problem since there is no arbitrary method for making this determination. In this instance, the order of the first observed maximum was obtained by varying the order number until the velocity calculated using Eqs. (9) and (11) appeared to be of the right order of magnitude. Table III lists the order numbers assigned in this way, together with the calculated acoustic velocities. An indication of the poor precision of the method is seen in the relatively wide variation in the calculated velocities.

There is some improvement, if one uses the graphical method indicated earlier. Since the positions of the maxima are known, the order numbers can be varied until the best approximation to a straight line is obtained. Figure 11 illustrates how this technique can be applied to the transmission curve (Fig. 12) of a $\frac{1}{8}$-inch copper plate. The square of the sine of the angle of incidence corresponding to the longitudinal interference maxima is plotted as a function of the square of the interference order. Figure 13 is an analogous plot for the shear waves. In Table IV, the indices of refraction obtained from the graphs in Figs. 11 and 13 are compared with those obtained by direct measurement of the angles of total reflection. The agreement is excellent.

In Table V, the elastic constants for a number of plastics are recorded. In thermoplastics the constants are low, indicating low resistance to deformation and flow. On the other hand, the thermosetting materials tend to have mechanical moduli approaching those of the glasses.

To investigate the proposition that the elastic constants should reflect changes in cure time of a thermosetting resin, a series of unfilled and cellulose filled melamine-formaldehyde resins were measured. As is well known, there is a pronounced change in the resistance of these resins to attack by boiling water as the cure time is varied. In Table VI, the elastic constants of several different resin samples of various cure times are recorded. Contrary to expectation, there is apparently no variation in the elastic constants with cure time. The only obvious explanation capable of reconciling these results with those obtained from such boil tests is that the boil test measures only an effective surface cure which is too thin to influence the bulk elastic constants.

The Effect of Current Magnitude upon the Behavior of a Superconducting Bolometer in Its Transition Region

NELSON FUSON
Department of Chemistry, Johns Hopkins University, Baltimore, Maryland
(Received September 14, 1948)

A simple theoretical analysis of the superconducting bolometer in transition is shown to be useful in predicting phenomena such as the shape of the transition curve and the dependence of the bolometer time constant upon the current magnitude. The analysis holds for bolometer currents smaller than a critical value which is equal to the square root of the ratio of the bolometer cooling constant to the bolometer dR/dT. For bolometers constructed in this laboratory this critical current value falls between 15 and 30 ma.

Qualitative extension of the theory to the case of large bolometer currents is made by considering the presence of regions at different temperatures within the bolometer and by summarizing the possible effects of the magnetic field due to the bolometer current upon the transition. Experimentally obtained transition curves are given for both small and large bolometer currents.

IN the course of a study of the infra-red sensitivity of superconducting bolometers,[1] a number of phenomena were observed which have particular interest in the general field of superconductivity. In order to introduce these effects in a proper setting a simple theoretical development will be given which will aid in predicting how various parameters affect the bolometer sensitivity, time constant, and shape of transition curve.

THEORY

The Uniform Temperature Bolometer

The bolometer potential, ΔV, for any bolometer in a series circuit, including a source of d.c. potential and a relatively large variable resistor, can be evaluated by the following expression:

$$\Delta V=\Delta(IR)=I\Delta R=I(dR/dT)\Delta T, \qquad (1)$$

where I is the direct current through the bolometer (throughout this paper I is assumed to be constant to the order of approximation required in this development), R is the bolometer resistance, dR/dT is the rate of change of bolometer resistance with temperature, and ΔT is the temperature change of the bolometer produced by changes in intensity of the radiant power incident upon the bolometer. From this relationship it is clear that for a given signal to produce a large potential drop across the bolometer the I, dR/dT and ΔT should all be as large as possible. The current, I, is an easy quantity to adjust while operating the bolometer, and it will be of interest a little later on in this paper to examine the limits within which it can be increased. dR/dT is a characteristic property of the material of which the bolometer is made rather than of its dimensions or construction. It depends on the method of preparation of the bolometer substance.[1,2]

Particular interest also centers upon ΔT. In order to find the dependence of this temperature fluctuation on the incident power fluctuation it is necessary to specify some conditions for the bolometer. Consider the special case of the superconducting bolometer.[3] Let the assumptions

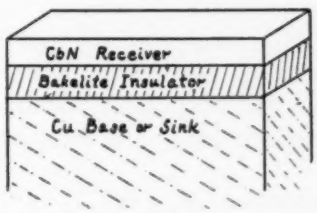


FIG. 1. Cross-section view of bolometer showing receiver, insulator, and sink.

[1] N. Fuson, J. Opt. Soc. Am. **38**, 845 (1948).

[2] R. M. Milton, Chem. Rev. **39**, 419 (1946), especially page 420 and reference 14 on page 433.

[3] The starting point for this discussion of superconducting bolometer theory is taken from reference 2. A somewhat different analytical approach is made by D. H. Andrews, Phys. Soc. Cambridge Conference Report, page 56 (1947). See also B. H. Billings, W. L. Hyde and E. E. Barr [J. Opt. Soc. Am. **37**, 123 (1947)] for a discussion of the theory of a thin "backed" room temperature bolometer. Their development differs from that given here since their quantity

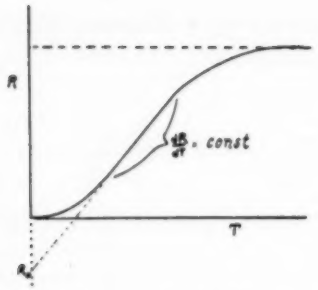


FIG. 2. Schematic diagram of transition curve, bolometer resistance *vs.* bolometer temperature.

be made that sine-wave modulated radiation falls uniformly on the bolometer receiver, that there are no appreciable temperature gradients in the bolometer, no radiation losses, no conduction losses through the leadwires, and that all cooling is by conduction through the electrically insulating "backing" layer separating the bolometer from the massive base which is its temperature sink (see Fig. 1). The law of conservation of energy for such a bolometer, expressed in terms of power, is

$$P-(1-\alpha)P+I^2R-k(T-T_s)=c(dT/dt), \quad (2)$$

where P is the power striking the bolometer, α is the absorption coefficient of the bolometer receiver, k is the heat leak or cooling constant of the insulating layer, T is the bolometer temperature, T_s is the temperature of the massive base or sink, and c is the thermal capacity of the bolometer. The cooling constant, k, is computed from the relation

$$k=\kappa a/1, \quad (3)$$

where κ is the coefficient of thermal conductivity of the backing layer, a is the area of the bolometer and 1 is the thickness of the backing layer. The thermal capacity, c, is computed from the relation

$$c=\sigma M/M_0, \quad (4)$$

where σ is the specific heat of the bolometer material, M is the mass of the bolometer, and M_0 is the molecular weight of the bolometer substance.

Introducing a change of variables, let $\theta=T-T_s$. Then, assuming that T_s is a constant or at most a slowly varying function of time, $d\theta/dT=dT/dt$, and Eq. (2) becomes

$$\alpha P+I^2R-k\theta=c(d\theta/dt). \quad (5)$$

R is a rather complex function of bolometer temperature (see Fig. 2) but if consideration is limited to the region of constant dR/dT, R can be expressed as a linear function of temperature

$$R=R_0+(dR/dT)T=R_0+(dR/dT)(T_s+\theta), \quad (6)$$

and, within this limited region, Eq. (5) becomes

$$c(d\theta/dt)+[k-I^2(dR/dT)]\theta = I^2[R_0+(dR/dT)T_s]-\alpha P. \quad (7)$$

Let the radiant power, P, be a sine-wave modulated function of the form

$$P=\Delta J(1-e^{i\omega t}), \quad (8)$$

where ω equals $2\pi f$, and f is the modulation frequency of the incident power (see Fig. 3). The complex number notation is used, P being understood to be the real part of the right hand expression in Eq. (8). If

$$Y=k-I^2(dR/dT), \quad (9)$$

and

$$Z=I^2[R_0+(dR/dT)T_s], \quad (10)$$

then, substituting Eqs. (8), (9), and (10) into Eq. (7) gives

$$c(d\theta/dt)+Y\theta=Z+\alpha\Delta J-\alpha\Delta Je^{i\omega t}. \quad (11)$$

A trial solution of Eq. (11), for $Y>0$, is

$$\theta=D+Ee^{-Yt/c}+Fe^{i\omega t}, \quad (12)$$

where D, E, and F must be determined. To do this, differentiate Eq. (12) to get

$$d\theta/dt=-(Y/c)Ee^{-Yt/c}+i\omega Fe^{i\omega t}. \quad (13)$$

Substituting Eqs. (12) and (13) into Eq. (11), and comparing coefficients, gives

$$D=(Z+\alpha\Delta J)/Y \quad (14)$$

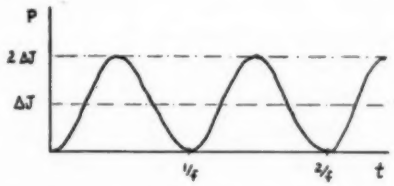


FIG. 3. Sine-wave modulated radiation signal incident on bolometer.

"$EI\alpha$" corresponding to "I^2dR/dT" in this paper was several orders of magnitude smaller than their cooling constant and could be neglected. This was not possible in the present development.

and

$$F=-\alpha\Delta J/(Y+i\omega c)$$
$$=-\alpha\Delta J(Y-i\omega c)/(Y^2+\omega^2c^2)$$
$$=-\alpha\Delta J(Y^2+\omega^2c^2)^{-\frac{1}{2}}e^{-i\varphi}, \quad (15)$$

where

$$\varphi=\tan^{-1}\omega c/Y. \quad (16)$$

The solution to Eq. (11) then is

$$\theta=(Z-\alpha\Delta J)/Y+Ee^{-Yt/c}$$
$$-\alpha\Delta J(Y^2+\omega^2c^2)^{-\frac{1}{2}}e^{i(\omega t-\varphi)}. \quad (17)$$

In the steady state as t becomes large the second term containing the unknown constant coefficient, E, approaches zero; the final result for θ in the steady state is the real part of the following expression:

$$\theta=(Z+\alpha\Delta J)/Y-\alpha\Delta J(Y^2+\omega^2c^2)^{-\frac{1}{2}}e^{i(\omega t-\varphi)}. \quad (18)$$

Thus θ has an equilibrium value (first term of Eq. (18)) modulated by an amount (coefficient of second term of Eq. (18)) whose phase angle, φ, is known. This solution is graphed in Fig. 4. The thermal periodic driving force is $\alpha\Delta J$ and the thermal impedance is $(Y^2+\omega^2c^2)^{-\frac{1}{2}}$. A mechanical analogy would be that of a mass, not attached to any spring, moving in one dimension in a resisting medium under the influence of a periodic driving force.

Any solution of Eq. (11) for $Y<0$ will include a positive exponential factor in t. Thus as t becomes large the temperature difference, θ, will increase without limit, i.e., the bolometer temperature, T, will fall outside the range of constant dR/dT, the range within which this development is valid. In order to have stability from this simple theoretical point of view, another limitation is thus added, namely, that $Y>0$, or, from Eq. (9), the bolometer current cannot exceed a certain critical maximum value, I_m, which is

$$I_m=[k/(dR/dT)]^{\frac{1}{2}}. \quad (19)$$

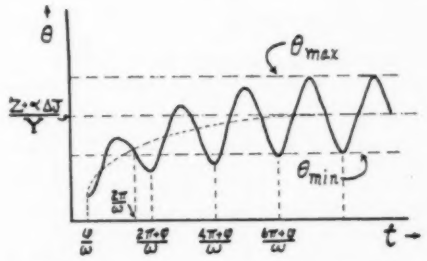


FIG. 4. Bolometer temperature *vs.* time, showing how the contribution of the exponential term disappears after a time equal to a few bolometer periods.

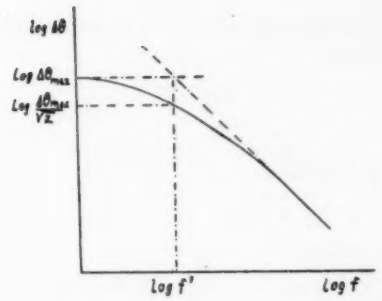


FIG. 5. Bolometer temperature fluctuation *vs.* modulation frequency plotted on log-log paper.

Figure 5 is a graph of the temperature fluctuation, $\Delta\theta=\theta_{\max}-\theta_{\min}$, computed from Eq. (18) and plotted as a function of modulation frequency on log-log paper. For slow modulation of the incident radiation ($f\to 0$, or $Y/c>\omega$) Eq. (18) reduces to

$$\theta=(Z+\alpha\Delta J)/Y-(\alpha\Delta J/Y)e^{i(\omega t-\varphi)} \quad (20)$$

and

$$\Delta\theta=2\alpha\Delta J/Y. \quad (21)$$

Similarly, for fast modulation ($f\to$large, or $Y/c<\omega$), Eq. (18) becomes

$$\theta=(Z+\alpha\Delta J)/Y-(\alpha\Delta J/\omega c)e^{i(\omega t-\varphi)}, \quad (22)$$

and

$$\Delta\theta=2\alpha\Delta J/\omega c. \quad (23)$$

These two asymptotic values of $\Delta\theta$, when drawn in on the log-log graph (Fig. 5), are straight lines, the former with zero slope, the latter with a slope of minus one. Their intersection defines a useful value of frequency, $\omega=\omega'$, for since

$$2\alpha\Delta J/Y=2\alpha\Delta J/c\omega', \quad (24)$$

then

$$f'=\omega'/2\pi=Y/2\pi c. \quad (25)$$

The value of temperature fluctuation calculated from Eq. (18) for this particular frequency is

$$\Delta\theta'=2^{\frac{1}{2}}\alpha\Delta J/Y=\Delta\theta_{\max}/2^{\frac{1}{2}}=0.707\Delta\theta_{\max}. \quad (26)$$

In terms of the bolometer time constant, τ, which for this system may be taken to be[4]

[4] This expression for τ can be indicated, non-rigorously, by pointing out that its reciprocal, Y/c, is the coefficient of t in the exponent of the second term in Eq. (18). If the bolometer is exposed to a sudden long pulse of radiant energy, this would be the dominant term controlling the rise of bolometer temperature.

TABLE I. Bolometer response at low frequencies.*

I ma	ΔV mv	Y $\frac{\mu\text{-wt}}{^\circ\text{K}}$	$\Delta VY/I$
3.5	0.023	1620	10.7
10.0	0.061	1400	8.7
33.0	0.250	—	
50.0	0.320	—	

* Bolometer 13-g1 is exposed to a constant radiant power modulated at 20 c.p.s. For currents listed in column 1, column 2 gives the corresponding bolometer outputs. The Y-values in column 3 are calculated from Eq. (9) (the last two values of I are greater than I_m so no Y could be computed). The numbers in column 4 should all be the same according to Eq. (29).

$$\tau = c/Y = c/(k - I^2 dR/dT), \tag{27}$$

Eq. (25) gives

$$\tau = 1/\omega' \tag{28}$$

which enables τ to be computed from experimental data,[5] even when neither the heat capacity nor the cooling constant of the bolometer is known.

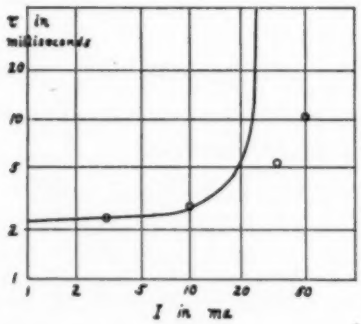


FIG. 6. Bolometer time constant *vs.* bolometer current. The solid curve is deduced from Eq. (27); the four individual points are the experimental results. (All for Bolometer 13-g1.)

[5] There are two possible ways of calculating ω' from Fig. 5. It may be got by taking the abscissa of either (a) the intersection of the two asymptotes, or (b) the point where the curve crosses the $\Delta\theta = 0.7\Delta\theta_{max}$ ordinate. These two ways give the same result when the curve is shaped according to theory as in Fig. 5. Experimentally the curve looks similar, but the slope of the fast modulation asymptote turns out to be nearer to -0.7 than to -1.0 (though it approaches somewhat closer to the theoretical value the larger the bolometer current!). The deviation of this slope from the predicted value may be associated with the fact that a truncated saw-tooth wave form was used rather than the sinusoidal wave form called for in the theory (see footnote 7 in reference 1). As a consequence of this deviation the values of τ obtained by method (a) are always larger than those obtained by method (b). Method (a) was used throughout this research. It has also been used in previous studies of superconducting bolometers (see references 1 and 2; see also E. E. Bell, R. F. Buhl, A. H. Nielsen, and H. H. Nielsen, J. Opt. Soc. Am. **36**, 355A (1946)).

Design Considerations

In the design of a superconducting bolometer to have a high sensitivity ratio (essentially $\Delta\theta/\Delta J$) when responding to a given modulation frequency, f, it can be seen either from Eqs. (21) and (23) or from the thermal impedance coefficient in Eq. (18) that α must be large and that Y and c must be made small. α is largest when the bolometer receiving surface is "black" to the radiation falling upon it;[6] c is minimized by making the dimensions of the bolometer small. The thinner the flake[7] and the smaller the receiving area the better, though the latter is usually specified by the application for which the bolometer is designed. In order for Y to be small (assuming for the moment that $Y \cong k$) a very thick layer of insulator is needed. However, since this also increases the time constant of the bolometer, k is usually adjusted to give the speed of response desired rather than being specified by sensitivity considerations. It should be noted here that the optimum frequency at which to operate a superconducting bolometer is f'. At higher frequencies sensitivity is lost, while at lower frequencies full advantage is not taken of the rapid response of the bolometer.

Modifications of the Theory for Large Bolometer Currents

For the more general case in which the simplifying assumptions of temperature uniformity made at the beginning of this section are no longer imposed, the quantitative analysis becomes exceedingly complicated and will not be attempted here. Yet it is clear that local irregularities in the bolometer flake[8] will introduce differential Joule heat production, there will be temperature gradients, and different regions of the bolometer flake

[6] For columbium nitride superconducting bolometers, experiments have shown that the natural CbN surface is superior as a radiation receiver to a CbN surface "blackened" with evaporated gold, platinum, etc. (See page 424–425 of reference 2.)

[7] Two different thicknesses of CbN flakes were used, 0.0254 mm and 0.0064 mm, respectively. The thinner flakes have proved somewhat superior in terms of sensitivity. Yet different samples of CbN of the same thickness differ so radically among themselves in sensitivity when made up into bolometers that it is difficult to assess the value of the thinner samples quantitatively without more extensive tests (see Table I in reference 1).

[8] Cracks, flaws, variations in thickness of the insulating layer and in the degree of nitridization of the strip, edge and end effects, etc.

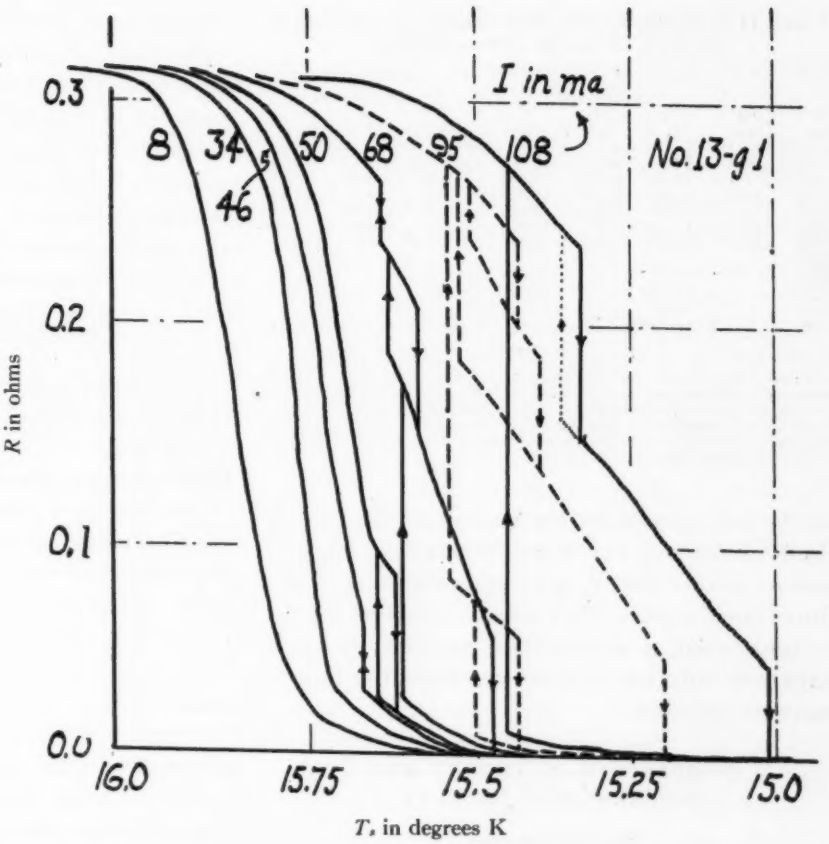


FIG. 7. Experimental CbN superconductor transition curves for bolometer No. 13-g1 taken for seven different currents. The $I=95$ ma curve is drawn with broken lines in order to eliminate confusion where curves for two different currents intersect. The dotted portion of the $I=108$-ma curve exists, but its precise location was not actually traced out by the potentiometer recorder at the time the rest of the transition was obtained.

will be at different temperatures. The temperature gradients, which may be neglected for low currents, will increase rapidly with an increase in the d.c. current in the bolometer. If temperature differences between local regions of the bolometer become great enough, it is probable that cooler regions will "go super" before the neighboring regions do during the transition cycle. After each such local region "goes super" the temperature distribution of the entire bolometer will suffer radical alteration.

One other consideration at high currents should be mentioned. Magnetic fields produced by the d.c. current in the superconducting bolometer will, if they are large enough, have influence upon both the location and the width of the temperature transition interval.[9] There is a wide variation in the magnitude and sign of these effects for different superconducting materials. The variation is particularly marked between superconducting metals and compounds and superconducting alloys. It is not yet clear just were columbium nitride fits into this range of behavior both because of the ambiguity as to whether or not it is a pure compound and because its critical magnetic fields, due to either an external field or

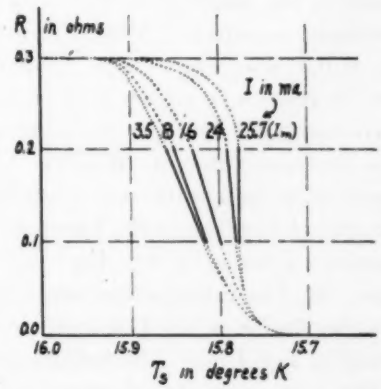


FIG. 8. Theoretical transition curves for CbN bolometer No. 13-g1, computed for five different currents including I_m. The dotted portions of the curve are drawn in qualitatively, the solid portions only being calculated from Eq. (31).

[9] See, for example, D. Schoenberg, *Superconductivity* (Cambridge University Press, Teddington, 1938), particularly the latter half of Chapter IV and all of Chapter VI.

TABLE II. Calculated value of bolometer temperature for different bolometer current values.*

		R=0.2 ohm			R=0.1 ohm		
I ma	I^2R μ-wt	T_s °K	ΔT °K	T °K	T_s °K	ΔT °K	T °K
0.0	0.0	(15.86)	0.0	15.860	(15.82)	0.0	15.820
3.5	1.8	15.86	0.001	15.861	15.82	0.0005	15.820
8.0	12.8	15.85	0.008	15.858	15.81	0.004	15.814
16.0	51.2	15.83	0.031	15.861	15.79	0.015	15.815
24.0	115.2	(15.80)	0.07	15.870	(15.77)	0.035	15.805
25.7	132.0	(15.79)	0.08	15.87	(15.76)	0.04	15.80
32.0		15.77			15.74		
46.0		15.73			15.69		
50.0		15.69			15.65		
68.0		15.62–15.58			15.60–15.52		

* Observed values of sink temperature (columns 3 and 6), calculated temperature differences between sink and bolometer (columns 4 and 7), and the resulting calculated values for bolometer temperature (columns 5 and 8) corresponding to different bolometer currents (column 1). The bracketed values of T_s were not observed directly but were extrapolated from experimental data.

to the field caused by the current in the sample itself,[9] have not yet been determined. One, of course, knows that if magnetic effects have any importance at all in CbN superconductors for the currents used, it will be most marked, and perhaps will only begin to occur, when the largest currents are used.

COMPARISON OF THEORY AND EXPERIMENTAL RESULTS

The Bolometer

The experimental results upon which this paper is based were observed qualitatively in a number of columbium nitride superconducting bolometers.[10] Most of the quantitative data obtained happened to be for bolometer No. 13-g1, and in the following discussion this particular bolometer will be used as the illustration. It is a bolometer of intermediate sensitivity. When "normal" its resistance is 0.3 ohm, when "super," 0.0 ohm; in transition its resistance region in which dR/dT remains constant lies between 0.2 ohm and 0.1 ohm. This constant value of dR/dT is 2.5 ohm/°K.[10a] Under optimum sensitivity conditions its time constant is 4.2 milliseconds. This bolometer's heat capacity, calculated by Eq. (4), is 3.8 microjoules/°K. The information required to obtain this value includes the bolometer receiving area, estimated as 6.4 mm², the bolometer thickness, 0.0254 mm, and the density, molecular weight and specific heat at mid-transition of columbium nitride which are 8.4 g/cc, 107 g and 0.30 joules/°K/g-mole,[11] respectively. Using this value of heat capacity and the value of the time constant with 3-ma current (see Fig. 6), Eq. (27) gives a value of 1650 microwatts/°K for the cooling constant.[12] Using this value of K and the dR/dT value mentioned above, Eq. (19) gives, for bolometer 13-g1, an I_m of 25.7 ma.

Comparison at Low Currents

The obvious place to check the theory is to see if the bolometer time constant is as current dependent as Eq. (27) would require. Figure 6 illustrates graphically both the theoretically predicted and the experimentally observed values of the time constant for different currents. For small current values the experimental points lie closely along the theoretical curve, but for values of I greater than $(2/3)I_m$ the divergence is very rapid. This is consistent with the limitation of the simple theory to small values of I.

Another experimental comparison with theoretical prediction can be made in terms of signal strength at low frequencies. From Eqs. (1) and (21) it follows that for a given bolometer under constant radiation conditions,

$$\Delta VY/I = 2\alpha\Delta J(dR/dT) = \text{constant}. \quad (29)$$

Table I shows in a limited way to what extent this appears to hold true. The data available are clearly not sufficient to be conclusive.

Still another comparison between theory and experiment is for transition curves obtained when the bolometer is shielded from radiation and when the bolometer temperature is changing very slowly with time. Under these conditions Eq. (2) reduces to

$$I^2R = k(T - T_s). \quad (30)$$

[10] All these bolometers, including the No. 13-g1 described below, are listed in the tables of reference 1.

[10a] This value is slightly larger than is given in reference 1, being based on additional data.

[11] This value was obtained in a private communication in March 1948 from Mr. George Armstrong, of the Johns Hopkins University Cryogeny Laboratory, who has been studying the thermodynamic properties of CbN at low temperatures.

[12] The cooling constant can also be calculated from Eq. (3), using 50 microwatts/°K/cm as the coefficient of thermal conductivity for Bakelite insulating layer at 15°K (see page 431 and footnote 7 in reference 2), and using the estimated values of bolometer area as 6.4 mm² and of Bakelite backing layer thickness as 0.03 mm. This gives 1100 microwatts/°K. Both calculated values of the cooling constant are highly approximate, the larger value being used arbitrarily in the calculations of this paper.

Since the sink temperature, T_s, is the directly measurable temperature[13] rather than the bolometer temperature, T, it is easy to obtain an experimental transition curve of R *vs.* T_s directly with a recording potentiometer monitoring the bolometer resistance for different constant bolometer currents as the bolometer is cooled through its transition region. Figure 7 is such a series of curves for bolometer No. 13-g1 using different values of I. It was compiled from pen-recorder chart records redrawn on a common baseline for comparison purposes. For the region of constant dR/dT, Eq. (30), when rewritten replacing T by $R/(dR/dT)$[14] and when solved for R, becomes

$$R=[k(dR/dT)/(k-I^2dR/dT)]T_s. \quad (31)$$

Figure 8 is a graph of Eq. (31) for five different values of bolometer current, including I_m, using values of k and dR/dT for bolometer No. 13-g1. The dotted portions of the curves are sketched in qualitatively since only the solid portions, which fall within the region of constant dR/dT, are calculable from Eq. (31).

A comparison of Figs. 7 and 8 shows qualitative agreement between experiment and theory in the smooth curves without breaks for small currents and in that after a certain current is reached (I_m in the theoretical case) there is a sharp vertical "break" in the curve. Table II is another way of comparing the experimental results with those predicted. In the upper half of this table column 3 contains the experimental values of T_s for different values of current when the bolometer resistance is maintained at 0.2 ohms. If to these are added the corresponding temperature differences, $T-T_s$, calculated from Eq. (30), the values of T shown in the fifth column are obtained. The closeness with which these T-values cluster around the T_s for zero (or very small) currents, in the case both for $R=0.2$ ohms and $R=0.1$ ohm, is also a measure of the agreement between theory and experiment.

[13] T_s is actually about 0.01°K greater than the experimentally measurable temperature of the hydrogen vapor. But this temperature difference remains constant regardless of variations in the temperature of the bolometer and so can be ignored unless absolute values of temperature are required.

[14] The assumption that dR/dT is independent of the magnitude of the bolometer current is assumed all along in this simple theory development.

Comparison at High Currents

A careful study of the observed transition curves in Fig. 7 for the larger currents will bring out the following facts: the sloping portions of the curves are reversible while the vertical "break" portions are irreversible except when they first appear (e.g., the "break" for $I=46$ ma and the top "break" for $I=68$ ma); the number, size and location of the "breaks" change as the bolometer current is increased; for the largest current used[15] the "cool-down" cycle of the curve is a complicated series of tiny slopes and "breaks," while the "warm-up" cycle is essentially one large "break" from zero resistance directly to normal resistance; the resistance of the sample for any given value of sink temperature may have any one of a number of values depending on previous history.

Two additional experimental facts, which cannot be seen in Fig. 7, may be summarized as follows: within the limits that the speed with which the transition cycle is traced out could be varied in this research, the speed had no effect upon the transition cycle (i.e., it did not change the position, size or number of the "breaks"); rough visual studies of voltage transients with an oscilloscope indicated that the time intervals during which the "breaks" occur are not all the same and that they certainly cannot be greater than a few milliseconds but are probably of much shorter duration.

Some observations can be added upon comparing these high current CbN superconductor transition curves with magnetization curves for iron rods in which "Barkhausen noises" are found.[16] The latter curves differ in that the details of the Barkhausen "breaks" (or the details of magnetic powder patterns in the case of powders) are *not* reproducible[17] as each cycle of the magnetization curve is made, and in that the sloping portions of the magnetization curves are

[15] Sixteen different transition curves for bolometer No. 13-g1, for currents ranging from 3 ma to 132 ma, have been recorded. Figure 7 shows only seven of these curves, the rest being omitted to avoid additional overlapping with the resultant confusion.

[16] R. M. Bozorth, Rev. Mod. Phys. **19**, 29–86 (1947), especially Fig. 34 on page 50; R. M. Bozorth and J. F. Dillinger, Phys. Rev. **35**, 733 (1930), Figs. 4 through 8; and H. J. Williams, Phys. Rev. **71**, 646 (1947).

[17] R. M. Bozorth, private communication of 3/19/48; also A. de Bretteville, Jr., Phys. Rev. **73**, 807 (1948).

not reversible, inasmuch as they are affected by hysteresis.

The remarks in the section on theory of transition at high currents will qualitatively explain some of the above phenomena, but the details await a more thorough analysis. It will be of interest to infra-red spectroscopists to note that no definite relationship exists between (a) the bolometer current which gives the best bolometer response when the latter is exposed to modulated infra-red radiation and (b) the bolometer current which brings the first "breaks." Usually they are rather close together, but the variation between different bolometers in this regard is great. The current at which the first "break" occurs usually sets an upper limit for current beyond which the bolometer, even though its sensitivity might increase, would be unstable and less adapted to use in radiation detection.

ACKNOWLEDGMENTS

This work has been supported by a grant from the Rockefeller Foundation. The author wishes to express his appreciation to Dr. Donald H. Andrews for many helpful discussions and for generous cooperation in the course of this research, and to the rest of the members of the Johns Hopkins University Cryogeny Laboratory for their interest and their aid in obtaining the experimental data.

The Effect of Uplift and Denudation on Underground Temperatures

A. E. BENFIELD

Department of Engineering Sciences and Applied Physics, Cruft Laboratory, Harvard University, Cambridge, Massachusetts

(Received July 9, 1948)

This paper gives calculations and computations, based on simple ideal conditions, of the possible disturbances of underground temperatures caused by uplift and denudation. The effects of the speed of denudation, the rate of uplift, and the time during which these two processes may have been operating, are illustrated by numerical tables applying to some specific ideal situations. The resulting disturbance of the heat flow in the earth (which is evaluated by making use of underground temperature gradients) is discussed. It is shown that uplift and denudation increase the heat flow above its equilibrium value, but that no very large increase is likely to occur unless the equilibrium heat flow is small and the speed of denudation is large.

INTRODUCTION

THE purpose of this paper is to show, by means of calculations, that underground temperatures are affected by the processes of uplift and denudation, such as operate in mountains and elevated regions. As the determination of heat flow in the earth depends on the measurement of underground temperatures,[1] heat flow will be correspondingly influenced by these processes. In general, the effects are not large, but it will be shown that in regions where uplift and denudation (particularly the latter) operate at a rapid rate, the effects may be appreciable. The effect of denudation has already been recognized and estimated, for instance, by Coster[2] in connection with his heat flow determinations for Persia, about which more will be said later.

The effect of *denudation* may be qualitatively seen by considering a plane area which is being uniformly and rapidly denuded at, say, a constant rate, and which is at the same time rising as a whole. Suppose for the moment that, while denudation is proceeding, the upward motion of the region is just large enough to keep the exposed denuded surface at such an elevation that its temperature is everywhere the same and does not vary with time. One can see that warm rocks are continually being brought up from below towards the exposed isothermal surface, which is in contact with the atmosphere. Therefore, the temperature gradient near this exposed surface will be greater than at depth, or in a region where denudation is proceeding slowly or not at

[1] A. E. Benfield, Proc. Roy. Soc. **A173**, 428–450 (1939).
[2] H. P. Coster, Monthly Notices of the Roy. Astr. Soc., Geoph. Suppl. **5**, 131 (1947).

all, and this gradient will be continually increasing with time.

In order to see qualitatively how *uplift* will further contribute to increase the temperature gradient, we may suppose that the upward movement is greater than we assumed above, so that the region suffering denudation is rising at such a rate that the exposed surface, as time goes on, is reaching into higher and colder levels of the earth's atmosphere. During this uplift, therefore, the temperature of the exposed surface of the rocks will not be constant, as assumed in the preceding paragraph, but will steadily fall;* this will have the effect of increasing the temperature gradient in the rocks near the surface, even if no denudation is occurring. Hence, the effects of denudation and uplift *add* in increasing underground temperature gradients. It turns out that, for the cases considered below, uplift does not produce a very large effect, but very rapid denudation, even if it operates over a relatively short length of time, can be of importance.

The matter is of some interest, because recently Birch[3] has again drawn attention to the fact that heat flow in mountainous regions should be larger than elsewhere for quite different reasons from those mentioned above, namely, because of the thickening of the relatively radioactive heat-producing granitic layer under mountains. Birch has succeeded in making a valuable heat flow determination for a high well in Colorado, and it seems likely that in the future more attention will be given to determining heat flow in elevated regions. It should be useful, therefore, to have an estimate of the effect of uplift and denudation in increasing heat flow.

In the present paper, in order to make calculations possible, we shall assume some simple ideal conditions, though not quite as simple as those assumed by Coster;[2] and we shall give tables of numerical results for the disturbed temperatures at various depths, which apply insofar as nature conforms to our ideal conditions. These various numerical tables of temperatures which follow are meant to serve as a means of comparing results, with a view to showing

TABLE I. Temperatures $T(x,t)$ in °C at various depths x for $t=0.3\times10^6$ years; $A=20$°C/km; $K=0.0130$ cm²/sec.

x(km)	$v=1\times10^{-9}$ cm/sec.* $M=0$°C	$v=1\times10^{-9}$ cm/sec.* $M=-10$°C†	$v=1\times10^{-9}$ cm/sec.* $M=-15$°C††	$v=3\times10^{-9}$ cm/sec.** $M=-10$°C	$v=10\times10^{-9}$ cm/sec.*** $M=-10$°C
0	0.0	0.0	0.0	0.0	0.0
1	20.54	23.40	24.83	24.57	29.02
2	40.95	45.96	48.47	47.99	55.53
3	61.25	67.86	71.17	70.49	80.14
4	81.47	89.21	93.08	92.26	103.30
5	101.62	110.17	114.44	113.49	125.43

* This rate is nearly 1 ft. per thousand years.
** This rate is nearly 3 ft. per thousand years.
*** This rate is nearly 10 ft. per thousand years.
† This corresponds to an uplift from sea level of about 2000 meters.
†† This corresponds to an uplift from sea level of about 3000 meters.

what factors are important in perturbing the temperatures underground. They are *not* correction tables, for in general the geology and tectonophysics of mountains and uplands is more complicated than has been here assumed for the purposes of calculation. Nevertheless, later in this paper we shall make an estimate of the possible disturbance caused by denudation in Persia, but one should bear in mind that it is based on our assumed simple ideal conditions.

An interesting idea, which suggests itself at this point, is that possibly the thermal effects of denudation and uplift on underground temperatures might someday be measured and used in order to calculate the rates of denudation and uplift in some given region. We shall not pursue this point of view here, however, as it seems quite unlikely that useful estimates can be made in this way. This is because of the many other factors which affect underground temperatures, and the difficulty of separating the various

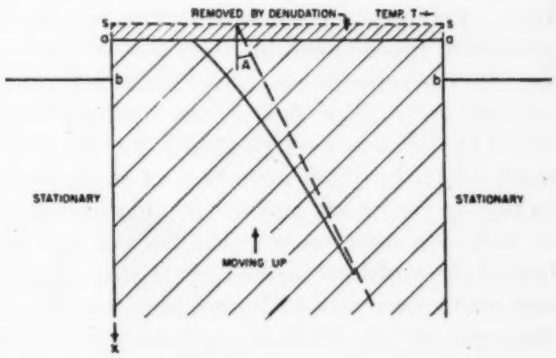


FIG. 1. The tableland at time t after the start of uplift and denudation. The dashed line (– – –) of slope A shows the supposed initial temperature distribution of constant gradient A, which existed before the start of uplift (when points s and b coincided). The solid line shows the resulting temperature distribution at the present time t.

* For the purposes of calculation we shall neglect the effect of secular or periodic climatic changes in the temperature of the earth's atmosphere.

[3] F. Birch, Trans. Amer. Geoph. Union **28**, 792–797 (1947).

TABLE II. Temperatures $T(x,t)$ in °C at various depths x for $t=1.0\times10^6$ years; $A=20$°C/km; $K=0.0130$ cm²/sec.

	$v=1\times10^{-9}$ cm/sec.			$v=3\times10^{-9}$ cm/sec.			$v=10\times10^{-9}$ cm/sec.		
x(km)	$\lambda=0$°C	$\lambda=-10$°C	$\lambda=-15$°C	$\lambda=0$°C	$\lambda=-10$°C	$\lambda=-15$°C	$\lambda=0$°C	$\lambda=-10$°C	$\lambda=-15$°C
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	21.06	22.73	23.57	23.29	25.03	25.90	32.52	34.50	35.50
2	41.97	45.08	46.63	46.10	49.32	50.93	62.82	66.44	68.24
3	62.74	67.08	69.25	68.48	72.95	75.19	91.24	96.19	98.66
4	83.40	88.79	91.48	90.47	96.00	98.77	118.07	124.11	127.12
5	103.95	110.22	113.35	112.14	118.54	121.75	143.58	150.49	153.94

disturbing effects. The problem would be rather similar to, and not easier than, that of trying to estimate from underground temperatures how the surface temperature of the earth has fluctuated during the recent ice ages. This has recently been carefully discussed by Birch.[4]

CONDITIONS AND CALCULATIONS

We shall now give the simple ideal conditions that will be assumed in the calculations. Some of these are shown graphically in Fig. 1. We shall consider an elevated plane tableland having a root of infinite depth and a horizontal extension *aa* (Fig. 1) large compared with its height *ab* above the surroundings, and we shall restrict our attention to points far from the edges of this tableland. In this way we reduce our mathematical problem to that of a moving semi-infinite medium. Furthermore, we shall assume that the material of the tableland is homogeneous throughout, that initially the surface of the tableland was at sea-level, and that initially the temperature gradient was everywhere a constant, A°C/cm. We shall call A the equilibrium gradient. (In Tables I–IV inclusive, we shall take $A=2\times10^{-4}$°C/cm$=20$°C/km which, in granite of thermal conductivity 0.0068 cal./cm-sec. °C, corresponds to a heat flow of 1.36×10^{-6} cal./cm²-sec.; these figures are close to those found by Birch in his Colorado well.[5] The tableland, which contains no source of heat, begins to rise at time $t=0$, and denudation also starts at $t=0$. We shall assume that the exposed surface of the tableland is maintained at a temperature which decreases with time (due to uplift) at the constant rate λ°C/sec., and that denudation occurs at a constant speed v cm/sec. The speed v, in other words, is the rate in cm/sec. at which the exposed plane surface, which is being constantly worn away, approaches a plane of reference embedded and fixed within the uplifting tableland. Hence, v may be called the speed of denudation; it is *not* the velocity of uplift of the tableland with respect to sea-level.

Our problem is then to find the temperature T at any time t and any depth x below the exposed surface of the tableland. The solution of this problem is of some mathematical interest, and has been discussed elsewhere.[6] The solution may be written

$$T(x,t)=T_0+A(x+vt)-\frac{1}{2}\left(A+\frac{\lambda}{v}\right)\times\left[(x+vt)erfc\left(\frac{x+vt}{2(Kt)^{\frac{1}{2}}}\right)-(x-vt)\times\exp\left(-\frac{vx}{K}\right)erfc\left(\frac{x-vt}{2(Kt)^{\frac{1}{2}}}\right)\right], \quad (1)$$

where K is the diffusivity of the homogeneous rock of the tableland, here taken as 0.00130 c.g.s., and

$$erfc\ y=1-erf\ y=1-\frac{2}{(\pi)^{\frac{1}{2}}}\int_0^y \exp(-u^2)du. \quad (2)$$

The constant T_0 is included to normalize the function $T(x,t)$ in order to make $T(x,t)=0$ when $x=0$ and $t=1948$ A.D. In other words, we are adjusting our computations to a mean annual surface temperature of 0°C on the exposed surface at the present time.

Tables I, II, III, IV and V give the values of the supposed present day temperatures in °C, calculated from Eq. (1), at depths of $x=0$, 1, 2, 3, 4 and 5 km below the exposed surface, for various different values of the parameters of Eq. (1).

[4] F. Birch, Am. J. Sci. (in press).
[5] F. Birch, Am. J. Sci. **245**, 733–753 (1947).
[6] A. E. Benfield, Quarterly of App. Math. (in the press).

TABLE III. Temperatures $T(x,t)$ in °C at various depths x for $t=3.0\times10^6$ years; $A=20$°C/km; $K=0.0130$ cm²/sec.

x(km)	$v=1\times10^{-9}$ cm/sec. $\lambda=0$°C	$\lambda=-10$°C	$\lambda=-15$°C	$v=3\times10^{-9}$ cm/sec. $\lambda=-10$°C	$v=10\times10^{-9}$ cm/sec. $\lambda=-10$°C
0	0.0	0.0	0.0	0.0	0.0
1	21.92	22.93	23.43	27.25	47.29
2	43.67	45.62	46.58	53.84	91.16
3	65.28	68.07	69.47	79.80	131.98
4	86.75	90.31	92.09	105.19	170.10
5	108.08	112.35	114.48	130.04	205.86

TABLE IV. Temperatures $T(x,t)$ in °C at various depths x for $t=10\times10^6$ years; $A=20$°C/km; $K=0.0130$ cm²/sec.

x(km)	$v=1\times10^{-9}$ cm/sec. $\lambda=0$°C	$\lambda=-10$°C	$\lambda=-15$°C	$v=3\times10^{-9}$ cm/sec. $\lambda=-10$°C
0	0.0	0.0	0.0	0.0
1	23.68	24.26	24.55	33.21
2	47.18	48.31	48.88	65.67
3	70.50	72.17	73.00	97.39
4	93.67	95.84	96.92	128.42
5	116.67	119.32	120.64	158.80

A comparison of these tables with one another brings out the following points:

1. A large disturbance of the temperature gradient from its equilibrium value A is chiefly caused by a very rapid speed of denudation v.

2. The percentage disturbance of the gradient is larger the smaller the value of the equilibrium gradient A. Hence, in evaluating the heat flow in regions where the measured temperature gradient is small, the effects of denudation and uplift may become relatively more important.**

3. When v is large the disturbance of the temperature gradient is largest for large values of t. This is because the denudation, which we suppose to operate at a constant speed v, has had a longer time in which to uncover rocks.

4. When v is small the magnitude of the disturbance of the temperature gradient does not depend sensitively on the value of t, unless A is small.

5. The temperature gradient is not seriously affected by any value of λ reasonable for, say, the Rocky Mountains, unless both t and A are small. This is because when t is large, denudation has been operating for a long time, and so the influence of λ is relatively small.

As has been mentioned above, Coster[2] has considered the effect of denudation on his heat flow values for Persia. He has evaluated the effect of the erosion of 1000 meters of rock in one million years. In order to make his estimate he assumed that the 1000 meters of rock were removed instantaneously 500,000 years ago, and that the surface of the ground stayed at the same level (which, in our notation, implies that $\lambda=0$). From this he has been able to estimate that, due to denudation, the increase in heat flow averaged over one million years is about 10 percent of the present value. Now, our Eq. (1) may also be used to make an estimate and, since it involves a gradual denudation rather than an instantaneous removal, it is perhaps in this respect more closely in accord with nature. If we assume, with Coster, that 1000 meters of rock have been removed in one million years, and use his constant $K=0.01$ cm²/sec., we have $v=3.17\times10^{-9}$ cm/sec., $t=10^6$ years, $\lambda=0$. If we further take A to be 15°C/km, we may use our Eq. (1), and find on doing so that the temperature at a depth of 1 km would be nearly 18°C greater than that of the surface. This estimate implies that the present heat flow is about 20 percent greater than the equilibrium, which is in good agreement with Coster's estimate of an average 10 percent increase over the whole period of one million years.

Coster, however, suggests that the denudation may have been proceeding for a few million years, rather than just for one. Referring again to Eq. (1), and setting $t=3\times10^6$ years while keeping all the other constants unchanged, we find that the temperature at a depth of 1 km would be about 20.6°C higher than at the surface, suggesting that the present heat flow may be some 35 to 40 percent greater than the equilibrium heat flow down a 15°C/km gradient.

It will be noticed, however, that Eq. (1) is not very suitable for making estimates of this kind. This is because the equilibrium flow A which appears implicitly in Eq. (1) is not, in fact, the measured quantity in heat flow work. A is really what we wish to find, after correcting

TABLE V. Temperatures $T(x,t)$ in °C at various depths x for $A=10$°C/km; $\lambda=-15$°C; $K=0.0130$ cm²/sec.

x(km)	$v=1\times10^{-9}$ cm/sec. $t=0.3\times10^6$ yr.	$t=1.0\times10^6$ yr.	$t=3.0\times10^6$ yr.	$t=10.0\times10^6$ yr.	$v=10\times10^{-9}$ cm/sec. $t=1.0\times10^6$ yr.
0	0.0	0.0	0.0	0.0	0.0
1	14.56	13.04	12.48	12.71	19.24
2	28.00	25.65	24.75	25.29	36.84
3	40.54	37.88	36.83	37.75	53.05
4	52.35	49.78	48.72	50.08	68.09
5	63.63	61.37	60.44	62.30	82.15

** It depends somewhat on the *cause* of A being low. For instance, if A is low because of a high thermal conductivity of the rocks, as in South Africa, the diffusivity K may also be high; and the higher the diffusivity, the less the temperature gradient near the surface will be disturbed, other things being equal, as reference to Eq. (1) or a consideration of the physical situation will show.

the measured gradient which is $\partial T(x,t)/\partial x$. Therefore, it is better to differentiate Eq. (1) with respect to x, and solve for A. On performing the differentiation we find

$$\frac{\partial T(x,t)}{\partial x}=A-\frac{1}{2}\left(A+\frac{\lambda}{v}\right)\left[erfc\left(\frac{x+vt}{2(Kt)^{\frac{1}{2}}}\right)-2v\left(\frac{t}{\pi K}\right)^{\frac{1}{2}}\exp\left[-\left(\frac{x+vt}{2(Kt)^{\frac{1}{2}}}\right)^{2}\right]\right.$$

$$\left.-\left(1-\frac{v}{K}\{x-vt\}\right)\exp(-vx/K)erfc\left(\frac{x-vt}{2(Kt)^{\frac{1}{2}}}\right)\right], \quad (3)$$

and if we let $\lambda=0$, we can write

$$A=\frac{\partial T(x,t)/\partial x}{1-\frac{1}{2}\left[erfc\left(\frac{x+vt}{2(Kt)^{\frac{1}{2}}}\right)-2v\left(\frac{t}{\pi K}\right)^{\frac{1}{2}}\exp\left(-\left(\frac{x+vt}{2(Kt)^{\frac{1}{2}}}\right)^{2}\right)-\left(1-\frac{v}{K}\{x-vt\}\right)\exp\left(-\frac{vx}{K}\right)erfc\left(\frac{x-vt}{2(Kt)^{\frac{1}{2}}}\right)\right]}. \quad (4)$$

If now we repeat our estimates for Persia, using Eq. (4) instead of Eq. (1) and let $\partial T(x,t)/\partial x = 15°C/km$ at the present time at a depth of 1 km, while keeping the other constants the same, we find that for 1000 meters of rock removed in one million years $A = 12.7°C/km$, and if the denudation has been proceeding for three million years $A = 11°C/km$, implying a disturbance of about 18 percent and 36 percent respectively, which are close to the estimates which we found above for these situations, using Eq. (1).

It may be of some interest to consider the effects of denudation in South Africa, where Krige[7] has measured some low temperature gradients of the order of 10°C/km, which have been used by Bullard[8] to combine with his conductivity measurements in order to compute heat flow. As mentioned above, a given rate of denudation may have a greater effect on the heat flow when the value of A is low than when it is high.** However, many of the rocks encountered in South Africa were apparently hard, and must weather slowly. Table V shows that no very startling disturbance of underground temperatures should be expected when $A = 10°C/km$ and v is small.

[7] L. J. Krige, Proc. Roy. Soc. **A173**, 450–474 (1939).
[8] E. C. Bullard, Proc. Roy. Soc. **A173**, 474–502 (1939).

CONCLUSIONS AND ACKNOWLEDGMENTS

Uplift and denudation have the effect of increasing the heat flow from the earth; this is due to the disturbance they cause to the temperature gradient underground. However, no very large percentage disturbance of the gradient is likely to occur unless the equilibrium heat flow is small and the speed of denudation v is large.

It is a pleasure to thank Francis Birch for some helpful discussions of this problem, and Ronold W. P. King for the loan of his calculator for making the numerical computations.

An Experimental Investigation of Formulas for the Prediction of Horn Radiator Patterns

G. A. Woonton, D. R. Hay, and E. L. Vogan
Department of Physics, University of Western Ontario, London, Ontario
(Received July 9, 1948)

The *E*-plane radiation pattern has been measured at short distances and both *E* and *H*-plane patterns at relatively great distances from the mouths of several electromagnetic horns. Measured values are compared with values calculated by the Kirchhoff formula and in addition the distant patterns are compared with values calculated by the corrected formula due to Stratton and Chu.

For the range of horn length and aperture considered, the Kirchhoff formula gives satisfactory agreement in the *E*-plane for angles as great as twenty degrees and at distances as short as one foot from the mouth of the horn; the *H*-plane predictions are not as satisfactory.

INTRODUCTION

THE prediction of the behavior of very short electromagnetic waves in physical situations, which would have been academic ten years ago, is now often of considerable importance. A problem of this kind, the prediction of the field in front of a radiating aperture, is fundamental to the design of microwave radiators and has become important in a growing number of laboratory applications. It has been recognized that this problem is a version of those optical diffraction problems which have been the subject of investigations for some hundreds of years and which, during the last forty years, have received several modified treatments which were undertaken in order to place the solution on a basis which is consistent with electromagnetic theory. The experiments and calculations which are reported in this paper constitute a part of a program of measurements which is planned to test the precision and convenience of the various formulae when they are applied to the prediction of fields produced by microwave radiators. The need for an analytical treatment of the error in these formulae expressed in terms of the wave-length of the radiation, the dimensions of the aperture, the phase and amplitude of the field in the aperture and in terms of the angular displacement of the radiator with respect to the receiver, is very apparent but the difficulties of an analytical treatment are such that at present there seems no way by which this information can be collected except through experiment.

Electromagnetic horns were made the instrument for testing diffraction relations at microwave-lengths only after some of the classical experiments of physical optics had been tried and had failed because of ground reflections and because of multiple paths between transmitter and receiver; the complete shielding of the horn ensures freedom from this last difficulty. The role of the various parts of the horn in these experiments is clearest when the horn is treated as a radiator but because of the theorem of reciprocity any conclusion concerning its performance which holds for it as a radiator can be applied to it as a receiver. Radiation issuing from the throat of the horn, if the discontinuity is not too great, passes down the body of the horn to the aperture as a spherical wave of such curvature that the electric lines can terminate normally on the flared sides of the horn. The radiation which floods the aperture appears to come from a source located at the intersection of the flared sides produced backward into the throat, so that the field in front of a horn radiator is the same as the diffraction field produced by a rectangular aperture in an infinite screen when illuminated by a point source. The problem differs from the common optical problem only in the wave-length of the radiation and the non-uniform intensity of the field in the aperture.

CALCULATION OF PATTERNS

A diffraction problem is solved by adding together in proper phase and intensity, at a field point, the radiation from every point on the boundary that is responsible for the phenomena; the applied mathematician achieves this result

by finding a solution to the inhomogeneous wave equation which reduces at the aperture to the field which is in fact maintained there by a primary source. A solution to this problem due to Kirchhoff[1] has long served as the relation from which most optical diffraction formulae are derived. Kirchhoff's solution usually is made in terms of a scalar function so that the predicted field is itself scalar but because of the special nature of the Laplace operator in a rectangular system of coordinates each of the rectangular components of a field vector can be treated as such a scalar function with the result that the three components of a vector field are predicted. A solution of the field of a radiator in terms of the Hertz vector has been given by Barrow and Greene.[2] The essential fallacy of the Kirchhoff solution is well known; Stratton and Chu[3] give a particularly clear discussion of the defects of the Kirchhoff solution as a preliminary to a corrected solution which they offer in the same paper. Kirchhoff assumed that an electromagnetic field, incident from behind on a hole in a screen, maintained in the hole its uninterrupted magnitude and polarization and was zero elsewhere; this assumption when applied to a metallic screen of good conductivity contradicts the field equations for, by it, electric lines are supposed to terminate discontinuously at the metal edge of the screen without the introduction of compensating effects. The solution due to Stratton and Chu permits the field to terminate in the same way but recognizes that the discontinuity can occur only if the electric lines terminate on alternating charge along the contour of the aperture; radiation from this alternating charge adds two contour integrals to the surface integral which alone constitutes Kirchhoff's solution. The Kirchhoff and the Stratton-Chu solution form the basis of the calculations which are discussed later.

A cylindrical system of co-ordinates (x, ρ, φ), chosen to agree with that of Chu[4] has been used in the calculations. The horn is assumed to be a sector of angular width φ_0 radians cut from a circular cylinder of radius ρ_1 and depth a. The horn is placed with its vertex along the x axis of the system so that the top and bottom surfaces

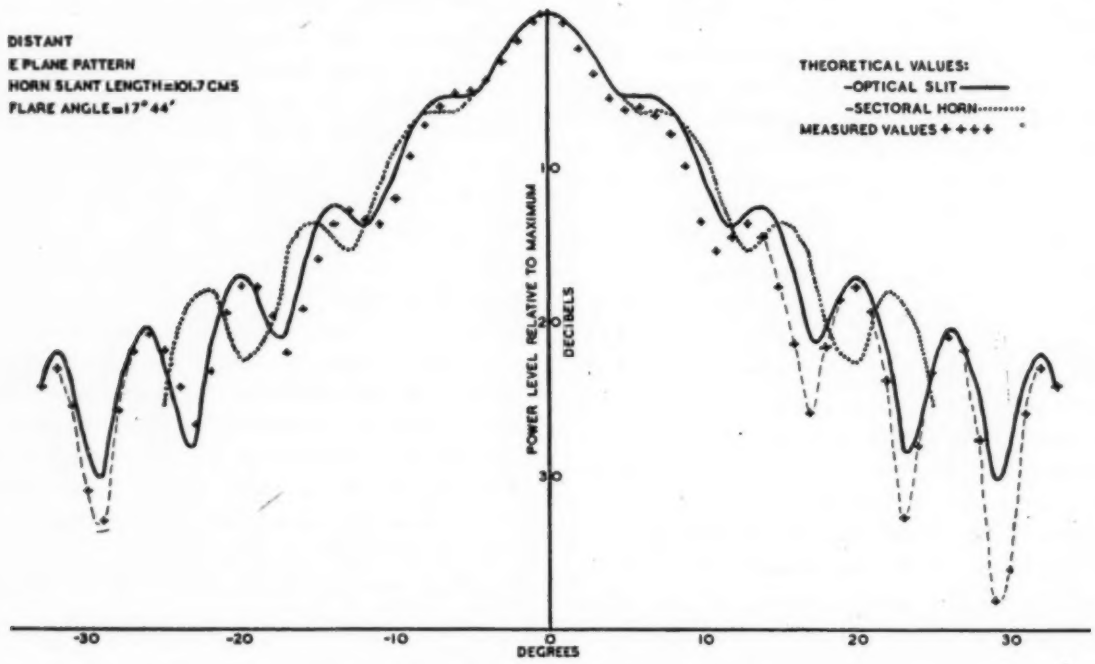


FIG. 1. Distant E-plane pattern: 100-cm horn.

[1] J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), 1st ed., p. 424.
[2] W. L. Barrow and F. M. Greene, Proc. I.R.E. **26**, 1498 (1938).
[3] J. A. Stratton and L. J. Chu, Phys. Rev. **56**, 99 (1939).
[4] L. J. Chu, J. App. Phys. **11**, 603 (1940).

are at $x=0$ and $x=a$ and the angular position of the horn with respect to the polar axis is such that the flared sides are at $\varphi=\pm\varphi_0/2$. The field point P, which in these calculations and experiments may be considered to be in the plane $x=a/2$ is designated by the coordinates $(a/2, R', \theta)$.

It is more convenient in experiment to rotate the horn around an axis defined by the intersection of the plane $\varphi=0$ with the aperture plane than it is to rotate it around the x axis as is contemplated in theory. All results both experimental and theoretical have been recorded in terms of the angular displacement α around the axis that lies in the aperture plane but for the purpose of calculation the coordinate θ is used and refers to rotation around the x axis in the manner customary to cylindrical systems. The two angles are related by

$$\tan\theta=\frac{\sin\alpha}{\rho_1(\cos\varphi_0/2)/R+\cos\alpha}$$

where R is the distance from the axis of rotation, used in experiment, to the field point P.

This correction is unnecessary in the calculation of distant patterns, since then R becomes great compared to the horn length.

Formulae, derived from the Stratton and Chu solution of the wave equation for the calculation of the distant patterns of electromagnetic horn have been published by Chu.[4] The data which in this paper is represented on the various graphs by a dotted line and which is designated "Sectoral Horn" in the legends was computed from formulae found in Chu's paper. These formulae have been calculated for a horn which flares in one plane only but for it two cases are presented in both of which the variable is θ: the pattern for rotation in the plane of the E vector is calculated for a case in which the electric lines are in the θ direction and the distant pattern E_θ is specified at the field point $(a/2, \infty, \theta)$; the pattern for rotation in the plane of the H vector is calculated for the same horn oriented in the same direction with respect to the coordinate system but in the aperture of which the electric lines point in the direction of the x axis so that the distant pattern is specified at the same field point in terms of E_x. Since the experimental arrangement is always such that the measured component of electric intensity is perpendicular to the radius vector, as are E_θ and E_x, it is sufficient to compute the square of the absolute value of the complex number E_θ

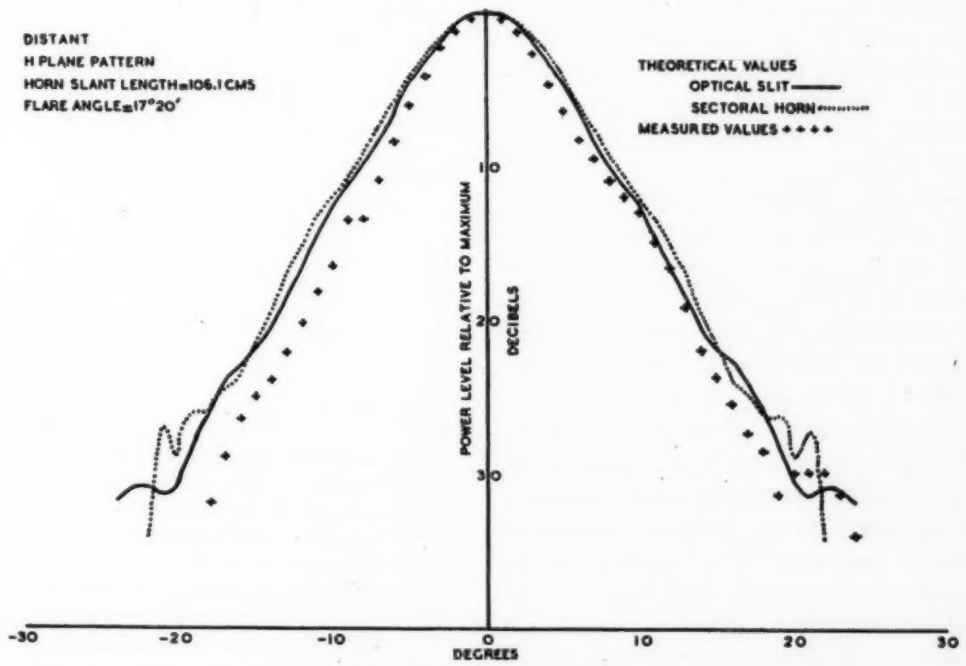


FIG. 2. Distant H-plane pattern: 100-cm horn.

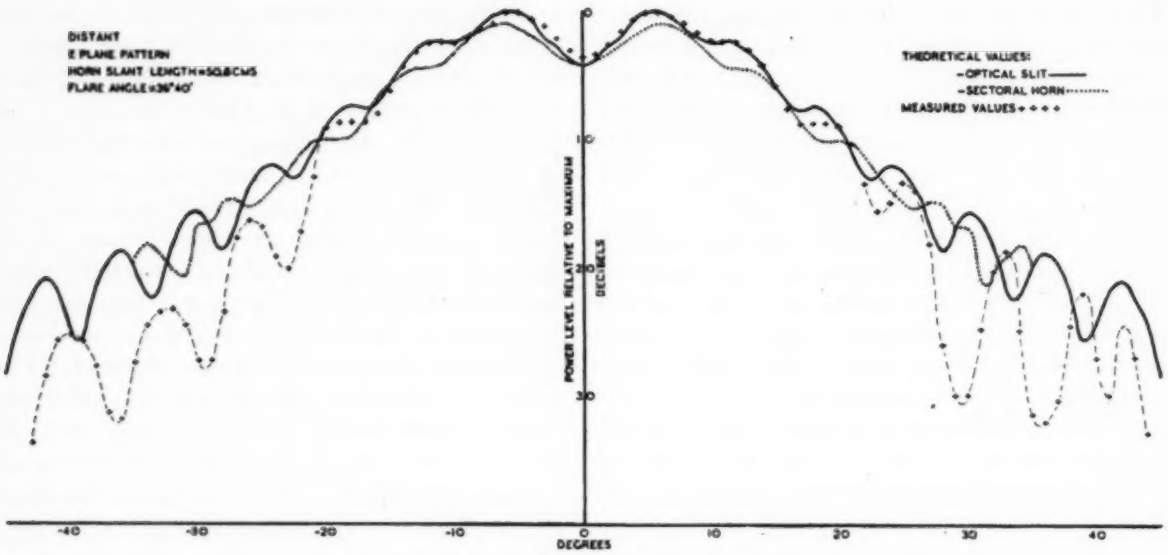


FIG. 3. Distant E-plane pattern: 50-cm horn.

for comparison with measurement in the E plane and of E_x for comparison in the H plane. The application of these calculations, which are computed for horns flared in one plane only, to horns flared in both the E and H plane introduces no new complications if relative rather than absolute values of power in the field are required, for, the mathematical statement of the pattern in one plane is independent of the dimensions and distribution in the other plane except for multipliers which are constant as long as rotation is in the specified plane; this principle is not unique but applies to the succeeding calculations as well.

All the common diffraction formulae of physical optics are derived from the Kirchhoff solution of the wave equation; the optical type of reasoning can be used to predict a scalar function V_p which determines the amplitude of the field at any distance R in front of an electromagnetic horn at a field point $(a/2, R', \theta)$ as θ is varied by rotating the horn in the plane of the electric vector. The horn is replaced in the imagination by a slit of the same dimensions as its aperture, which is illuminated by uniform cylindrical waves from a line source. The radius of the cylindrical wave front is taken to be the apparent radial length of the horn ρ_1 where ρ_1 is measured from the point of intersection of the flared sides produced into the throat. (Since the sinusoidal taper of the electric field amplitude, imposed by the requirement of zero tangential field at the boundary, is all in the x direction, the use of a uniform cylindrical wave to illumine the slit is permissible.) The scalar function which is predicted in this way is stated by the following relation:

$$V_p = \Re e \frac{jA}{2\lambda\rho_1 b} \exp\left(j\left[\omega t - \frac{2\pi}{\lambda}(\rho_1 + b)\right]\right)(1+\cos\theta)$$

$$\times \frac{2a}{\pi}\left(\frac{\lambda}{2}\cdot\frac{\rho_1 b}{\rho_1 + b}\right)^{\frac{1}{2}} \int_{\nu_9}^{\nu_{10}} \exp(-j(\pi\nu^2/2))d\nu$$

where

$$\nu_{10} = \left(\frac{2\rho_1}{\lambda}\cdot\frac{\rho_1+b}{b}\right)^{\frac{1}{2}}\left(\frac{\varphi_0}{2} - \theta\right)$$

$$\nu_9 = -\left(\frac{2\rho_1}{\lambda}\cdot\frac{\rho_1+b}{b}\right)^{\frac{1}{2}}\left(\frac{\varphi_0}{2} + \theta\right),$$

and in which

A = the amplitude of the electric field, assumed uniform;
ω = the angular velocity of the field;
λ = wave-length of the radiation in the horn;
ρ_1 = the apparent radial length of the horn;
φ_0 = the flare angle of the horn;
a = the dimension of the horn in the unflared direction;
θ = the angle between the axis of the horn and the line joining the vertex to the field point P;
b = the distance from the optical pole of the system to the field point $P(a/2, R', \theta)$.

For very distant field points, because b is great compared to ρ_1, this relation can be simplified by the use of the approximations:

$$\frac{\rho_1+b}{b}\doteqdot 1 \quad \text{and} \quad \frac{\rho_1 b}{\rho_1+b}\doteqdot \rho_1.$$

This formula for the calculation of the distant pattern is an improvement over the relation which was previously used by Woonton and Tillotson:[5] greater accuracy results from the use of the true flare angle, φ_0, instead of an approximation for it and in the explicit statement of the inclination factor $(1+\cos\theta)$; the common constant approximation for $(1+\cos\theta)$ is not justified in experiments in which the horn is rotated.

As in physical optics it has been assumed that the square of the absolute value of the scalar function V_p is proportional to the power in the field and this square has been used to obtain the data designated "optical slit," which is represented by the dashed line in the various graphs. In the absence of any other information on polarization and because the predicted values agree with measurement it is to be assumed that E_θ is proportional to V_p.

The same H plane but different E plane patterns should be predicted for an electromagnetic horn by the Kirchhoff formula and by Stratton and Chu's modification of it. It will be recalled that the difference between the two solutions is due to the line of alternating charge on which Stratton and Chu assume the electric lines terminate. Because the tangential component of electric intensity must vanish at the metal boundaries of the horn there is no distribution of radiating charge along the two edges of the horn aperture that are parallel to the E vector; charges of equal magnitude radiating in opposite phase are distributed along the other two edges. The symmetry of the system is of such nature that radiation of equal intensity but opposite phase arrives from these two sets of charge at a field point in the H plane and there cancels provided only that the horn is rotated in the plane of the magnetic vector and for this reason the same H plane prediction should result from formulae derived from either the optical or the electromagnetic relation.

The derivation of a horn pattern formula for rotation in the plane of the H vector is more difficult than the E plane calculation because of

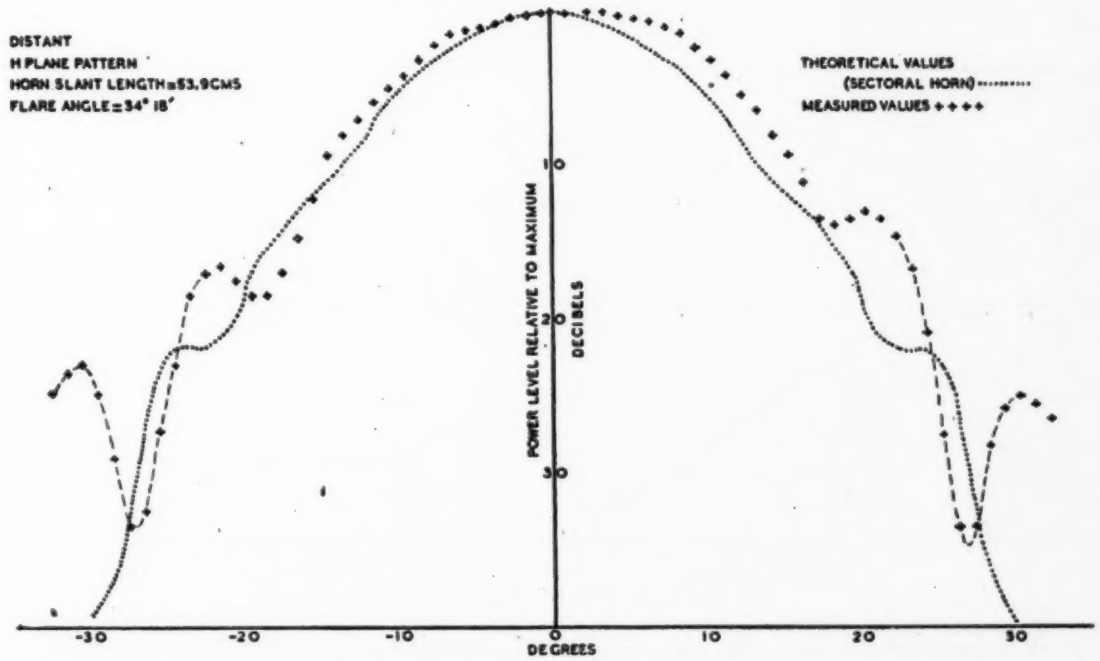


FIG. 4. Distant H-plane pattern: 50-cm horn.

[5] G. A. Woonton and J. G. Tillotson, Can. J. Research 25, 315 (1947).

the sinusoidal variation in intensity of the field in the aperture in the H direction. It is possible to derive a formula in the same manner as for the E plane pattern but the derivation leads to integrals which are very complicated. The relation on which the calculations were based which are designated in this report as H-plane "optical-slit" calculations was first discussed[6] and given experimental verification[7] by Michelson and has been put into the present useful form by Spencer.[8–11] Michelson observed that in optical diffraction the relation between the function of the direction cosines that describes the Fraunhofer pattern and the function of the coordinates that describes the field in the aperture is the same as the relation between the functions of a Fourier transform pair. Spencer's operational methods for finding, in series form, the Fourier transform of the aperature field distribution function and the numerical tables of transforms that he has compiled makes it possible to apply the method to the calculation of the distant H-plane patterns of horn radiators.

The Fourier transform method applies in all cases where the diffraction pattern can be linked to the aperture distribution function by a Fourier integral. Michelson observed the connection first in relation to a Fraunhofer pattern but symbolically at least it can be made to hold for any pattern Fraunhofer or Fresnel by the use of complex aperture distribution functions, provided only that the Kirchhoff formula describes the situation with sufficient accuracy. The experimental results which follow suggest that transform methods are adequate for all distant patterns both E-plane and H-plane and will very probably give excellent near pattern results.

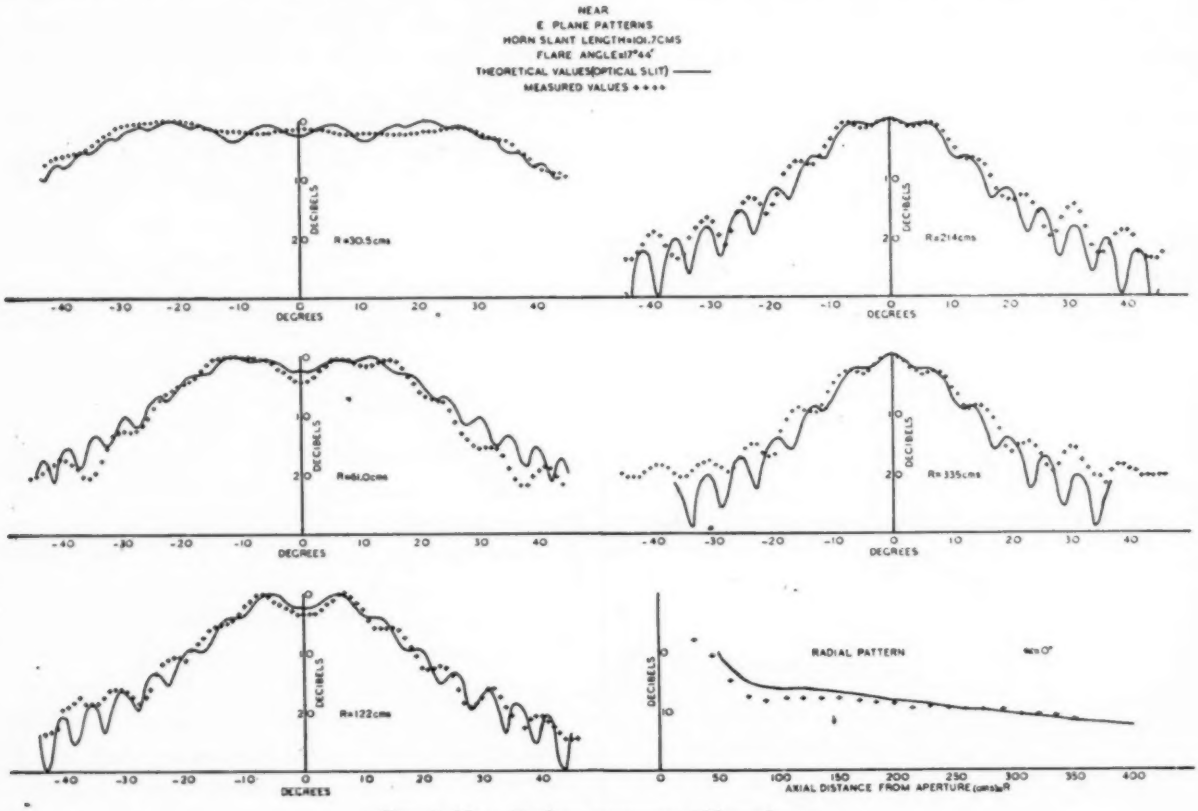


FIG. 5. Near E-plane patterns: 100-cm horn.

[6] A. A. Michelson, Phil. Mag. 9, 506 (1905).
[7] A. A. Michelson, *Studies in Optics* (The University of Chicago Press, Chicago, 1927), 1st ed., p. 60.
[8] R. C. Spencer, Radiation Laboratory Report T-7, Oct. 1942.
[9] R. C. Spencer, Radiation Laboratory Report S-58, Aug. 1945.
[10] R. C. Spencer, Radiation Laboratory Report 762–1, Jan. 1946.
[11] R. C. Spencer, Radiation Laboratory Report 762–2, Mar. 1946.

EXPERIMENTAL METHODS AND RESULTS

The copper horns on which experimental measurements were made were flared linearly in both the E and H plane from the end of a piece of standard, laboratory, X-band wave guide ($\lambda = 3.2$ cm) to a square aperture 32 cm to the side. Two horns were measured to obtain the data which is presented in the various figures. These horns were approximately 100-cm and 50-cm long but because of the slightly different flares in the two planes, made necessary by the rectangular throat, the flare angles and slant lengths in the two planes are slightly different; the independent treatment of the two planes was discussed in an earlier paragraph.

The distant patterns Figs. 1, 2, 3, 4 were measured by rotating the horns as receivers on a turntable about 15 feet above the ground in an open space which was free of obstruction at a distance of 88.5 m from a source of 3.2 cm waves. The experimental arrangements were as described by Woonton and Tillotson[5] except that the tuned amplifier and its stepped volume control were calibrated in advance of the experiment and the attenuator box eliminated from the circuit.

The near pattern measurements, Figs. 5 and 6, were made indoors with the horn as the transmitter and the open end of a wave-guide crystal mixer section as the receiver. The horn was mounted on a turntable about three feet above the top of a laboratory table and the receiver at the same height, on the end of a wooden rod mounted in the carriage of an optical bench. No precautions were taken against reflections except to mount the apparatus away from the wall and to remove large reflecting objects from the immediate vicinity and no troubles were experienced with reflections. The apparatus and methods for near pattern measurements, except for the rearrangements which have just been described were the same as for the distant pattern measurements.

The measurements, calculations and conclusions are not as clear cut for the H-plane patterns (Figs. 2 and 4) as for the E-plane. The close

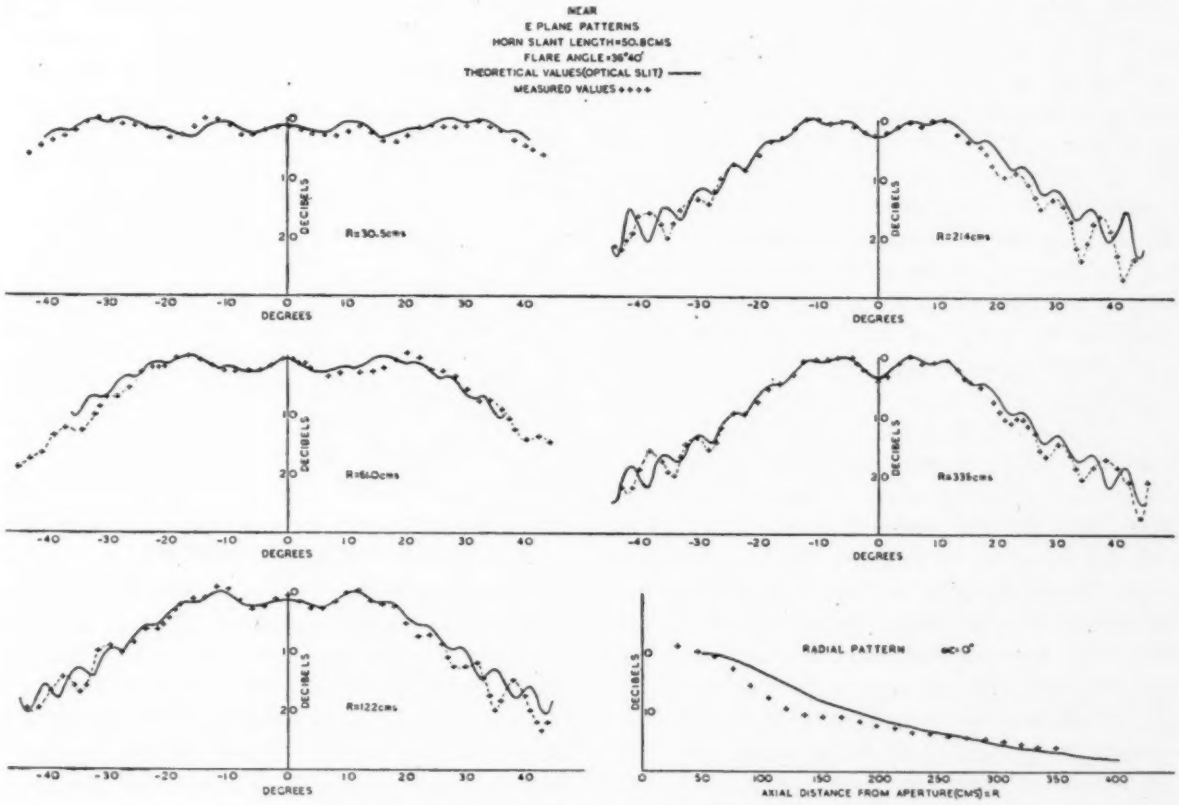


FIG. 6. Near E-plane pattern: 50-cm horn.

agreement in Fig. 2 between optical slit and sectoral horn calculations is interesting but not unexpected; in fact, the discrepancies that exist can only be due to differences in the approximations. The good agreement between the several calculations and their rough agreement with measurement suggests that there are no fundamental errors present in interpretation or computation but the rough agreement between prediction and measurement in Fig. 2 and the very poor agreement in Fig. 4 suggest that more measurements and calculations are required before it can be said that the *H*-plane patterns of these radiators can be calculated with precision; the error may lie in the assumed sine form of the aperture field taper. Numerical tables of Fourier transforms are not extensive enough to permit the calculation of optical slit data for Fig. 4. Even though the same prediction should result from the sectoral horn calculation, the transform method is so much less cumbersome and so much less time consuming that a considerable expenditure of effort is justified in the extension of existing numerical transform tables.

Distant *E*-plane patterns for the two horns have been predicted by both sectoral horn and optical slit formulae (Figs. 1 and 3) and near patterns have been predicted for the two horns, at distances of 1, 2, 4, 7 and 11 feet (Figs. 5 and 6) by the optical slit formula which alone is applicable. No very definite conclusions can be drawn from the comparisons presented by Fig. 3 for the distant pattern of the 50 cm horn partly because of the double hump which leaves some choice in the selection of the zero decibel level. The curves, adjusted as in Fig. 3, indicate that the optical slit is definitely more accurate than the sectoral horn formula for angular displacements less than twenty degrees and much less accurate for greater displacements. It has not been possible to shift the level of the sectoral horn calculation so as to improve the agreement at small displacements without destroying the agreement at wide angles. Figure 1, which compares the various computations and measurements for the 100 cm horn is very definite. The optical slit relation is no less accurate than the sectoral horn formula at small angles and very much more accurate at wide angles; the agreement between measurement and optical slit prediction at angles as great as thirty-five degrees is as remarkable as is the disagreement with the sectoral horn calculations.

Figures 5 and 6, the near patterns, show that it is quite possible to predict the field over a wide range of angle, for any distance greater than a foot from the mouth of these horns by purely optical relations. The level of any pattern relative to the level of another pattern at a different distance is in some doubt as shown by the radial patterns of Figs. 5 and 6; it is possible that this discrepancy is caused by reflections from the top of the table that supported the apparatus. No explanation can be offered for the fact that although the distant pattern predictions are more accurate for the longer horn, the agreement in these near patterns is better with the shorter horn.

CONCLUSIONS

Calculations based on relations taken from physical optics can be used to predict the magnitude of the *E*-plane field of a microwave horn radiator with good accuracy and the *H*-plane field with fair accuracy if the radiator is ten wave-lengths across; prediction may be possible with the same accuracy for much smaller radiators but no tests have been made. In the *E*-plane it has been shown that the optical predictions are accurate to within roughly one decibel up to angles of twenty degrees and down to distances of one foot from these radiators. The labor of calculation with optical relations is much less than with the corrected diffraction relations and except in the matter of polarization the accuracy is at least as great. Further measurements and calculations are required, with radiators of other dimensions and other aperture field distributions in order to determine the source of the discrepancies which have been observed.

ACKNOWLEDGMENT

This research was made possible by a grant from the National Research Council for the purchase of apparatus. Two of us, D.R.H. and E.L.V. wish to thank the Ontario Research Commission for studentships granted to carry on this work. The authors thank Mr. D. G. Rumbold for his constant and skillful technical assistance.

The Incremental Friction Coefficient—a Non-Hydrodynamic Component of Boundary Lubrication

J. T. Burwell and C. D. Strang
Lubrication Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 19, 1948)

In investigations of lubricated sliding friction between two crossed cylinders it is found that for hard, smooth surfaces the friction force is a straight-line function of the applied load, although this line does not generally pass through the origin. The slope of this line defines a friction coefficient, designated as the incremental friction f_Δ, which is thought to have certain advantages in the study of non-fluid lubricated friction. It is obviously independent of the load. It is found to be independent of viscous fluid effects and hence permits investigation of non-fluid friction at reasonable speeds. In spite of being devoid of any viscous component, it is found in a great many cases to decrease with increasing speed, U, over at least part of the range, and this variation often appears to be a linear function of the logarithm of the speed. The curve of f_Δ *vs.* ln U seems to be characteristic of the chemical structure of the lubricant and of the rubbing surfaces. It is not contended that viscous effects are absent under the conditions of these experiments but simply that the value of f_Δ, which is not the *total* friction coefficient, does not, because of its method of determination, include such effects.

INTRODUCTION

One of Coulomb's postulates states that the coefficient of friction is largely independent of the rubbing velocity over a wide range of speed. This has been generally taken to apply to boundary friction as well as dry friction, since in relatively few cases is friction actually "dry." A number of workers[1–3] have confirmed this conclusion experimentally for boundary friction while others[4–6] have found systematic variations, generally consisting of a decrease in friction coefficient with increasing speed. This latter result has led to the conclusion[7] that in boundary friction experiments, even under the most disadvantageous geometry, there is always a certain amount of hydrodynamic friction present, especially when an excess of lubricant is present on the rubbing surfaces. This conclusion is based on the assumption that an increase in speed produces a partial transition from boundary lubrication with a friction coefficient of about 0.1 to full-fluid or hydrodynamic lubrication with a friction coefficient of about 0.005. (However, once in the completely fluid region, increasing speed can only increase the friction coefficient.) Other researchers[8–10] completely avoid boundary friction experimentation at moderate speeds, and limit their rubbing velocities to very low values ranging from 0.003 to 0.01 centimeter per second. This lessens the possibility of fluid lubrication, but at these slow speeds relaxation oscillations or "stick-slip" phenomena may become evident if the static friction coefficient is larger than the kinetic[11,12]. This tends, in certain instances, to obscure the true kinetic friction value at the speed in question. Furthermore, most practical boundary-lubricated machinery must operate at moderate speeds, and information obtained at extremely low rubbing velocities may not be entirely applicable.

It is the purpose of the present article to describe an experimental procedure whereby it is believed that a component of the friction can be isolated from all hydrodynamic effects and determined quantitatively over a wide range of rubbing velocity. Some results will then be given showing how the friction coefficient, determined in this way, varies with speed.

[1] W. G. Beare and F. P. Bowden, Phil. Trans. **A234**, 329 (1935).
[2] R. L. Johnson, M. A. Swikert, and E. E. Bisson, N.A.C.A. Tech. Note 1442 (Oct. 1947).
[3] T. Sasaki, Bull. Chem. Soc. Japan **13**, 134 (1938).
[4] R. F. Wilson and D. P. Barnard, Ind. and Eng. Chem. **14**, 683 (1922).
[5] P. G. Forrester, Proc. Roy. Soc. **A187**, 88 (1947).
[6] J. R. Bristow, Proc. Roy. Soc. **A189**, 88 (1947).
[7] P. G. Forrester, J. Inst. Metals **73**, 573 (1947).
[8] F. P. Bowden and L. Leben, Phil. Trans. **A239**, 1 (1940).
[9] J. L. Heaton, J. R. Bristow, G. Whittingham, and T. P. Hughes, Nature **150**, 520 (1942).
[10] J. J. Frewing, Proc. Roy. Soc. **A181**, 23 (1942)
[11] H. Blok, J. Soc. Auto Eng. **46**, 54 (1940).
[12] F. Morgan, M. Muskat, and D. W. Reed, J. App. Phys. **12**, 743 (1941).

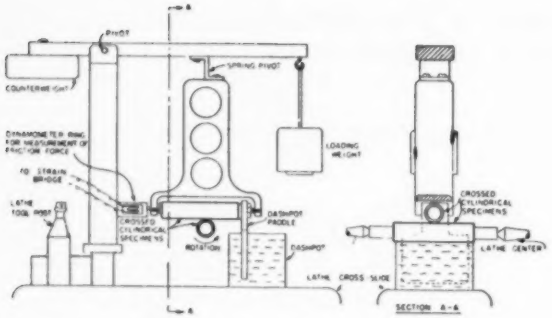


FIG. 1. Schematic view of friction apparatus.

APPARATUS

Factors governing the geometry of specimens to be used in boundary friction experiments have been discussed in detail by Holm,[13] Bowden and Tabor,[14] and others. They have shown that since two nominally flat surfaces actually touch at only a few small isolated areas, it is preferable to localize the true area of contact by using surfaces which are convex in the vicinity of contact. Therefore, in order (a) to achieve a concentrated and easily located area of contact, (b) to furnish a geometry unfavorable to the formation of hydrodynamic lubricant films, and (c) to facilitate the production of reproducible surface finishes, a geometry consisting of cylindrical friction specimens crossed at right angles was chosen for this investigation. The rubbing is accomplished by rotating one cylinder about its own axis while the other is held stationary.

The friction apparatus employed is shown diagramatically in Fig. 1. A heavy lathe bed serves as a rigid foundation for the apparatus and provides a source of rotation for the moving specimen. The stationary specimen is carried by a support which is mounted on the lathe carriage. It may be kept stationary so that it will run repeatedly over the same track on the rotating specimen, or the lead screw on the lathe carriage may be engaged to make the stationary specimen move axially along the rotating specimen, thus always running over a fresh track in the form of a spiral.

The support for the stationary specimen consists of a rigid upright column, mounted in the toolholder, at the top of which is pivoted a hroizontal loading arm. To this arm is spring-pivoted a rigid vertical member having a yoke at its lower end in which the stationary specimen is held on a removable mandrel. The load exerted between the crossed cylindrical specimens is supplied by weights suspended from the end of the horizontal loading arm. A counterweight at the other end of the loading arm counterbalances all of the assembly except the loading weight.

The rotating specimen is mounted on a mandrel which, in turn, is mounted on the centers of the lathe. The turning moment is furnished by the lathe faceplate and driving dog.

Friction between the crossed specimens tends to drive the yoke supporting the stationary specimen toward the upright column, and this motion is restrained by a dynamometer ring inserted between the yoke and the upright column. Deflection of the dynamometer ring by the friction force results in compressive strain at two points on the inner surface of the dynamometer ring and tensile strain on the outer surfaces. Four electrical resistance wire strain gauges (Baldwin type SR-4) are mounted on the inner and outer surfaces of the dynamometer ring at these points and form the four arms of a Wheatstone bridge circuit in such a sense that all resistance changes caused by strain in the ring add to increase the sensitivity of force measurement. On the other hand, changes in resistance resulting from temperature and vertical shearing forces cancel. As shown in Fig. 2, the bridge is balanced by means of a calibrated slide wire, using a Leeds and Northrup type P galvanometer having a time constant of eight seconds as the null indicator. A protective resistance in series with the galvanometer can be bypassed by means of a toggle switch to increase the sensitivity for final balancing. A variable resistor permits the zero position to be set, and both it and the slide-wire are shunted for greater sensitivity. Incipient vibration of the yoke, induced by relaxation oscillations of the rubbing specimens, is damped by a simple dashpot assembly.

A number of mechanical tests were made on the apparatus to insure against the possibility of structural deflections causing spurious friction readings. These were made statically, without rotation of the specimens. The friction force in

[13] R. Holm, *Electric Contacts* (H. Geber, Stockholm, 1946).
[14] F. P. Bowden and D. Tabor, Proc. Roy. Soc. **A169**, 391 (1939).

the pivoted joint between the upright column and the loading arm was found to be less than one-half of one percent of the lightest load applied in these experiments. Bending resistance of the spring pivot between the loading arm and yoke was canceled out in the calibration of the force measuring system. This was done by exerting known horizontal forces on the stationary specimen at the operating level and reading the corresponding values indicated on the calibrated slide wire. The calibration curve was found to be a straight line with load, and the over-all sensitivity of the measuring system was found to be 0.5 gram in the friction force. A check was made to determine the effect on the dynamometer ring readings of changing the specimen loading weights. This was done by preloading the yoke against the dynamometer ring with a known force, with the friction specimens in place but not rotating, and then varying the static load on the specimens. This showed that the force exerted on the dynamometer ring was completely independent of the normal load. Similarly, dial-indicator measurements on the spring pivot between the loading arm and the yoke showed no lateral deflection for the maximum vertical load applied. A dial indicator was used to determine the spring constant of the dynamometer ring, and it was found to be 45,400 grams per centimeter. Hence, for the friction forces involved in the boundary lubrication experiments, the maximum horizontal translation of the stationary specimen is less than 0.0025 centimeter. This is made possible by the great sensitivity of the strain gauge bridge. The lathe spindle is supported in plain bearings and is belt-driven from a separate motor and countershaft system which permits a choice of thirteen spindle speeds ranging from 0.015 revolution per second to 11 revolutions per second. With the 2.181-cm diameter specimens this corresponds to a linear speed range of 0.100 to 75.7 centimeters per second.

MATERIALS AND EXPERIMENTAL PROCEDURE

The friction specimens used in the present investigation consist of two standard automobile engine piston pins 2.181 cm in diameter and 7 cm long. They are made of SAE 1117 steel, case carburized and hardened to a value of 91 on the Rockwell 15-N scale. They are superfinished to a surface roughness of 1 or 2 micro-inches r.m.s., as determined by an Abbott Profilometer. Following Bristow,[6] they are cleaned by scrubbing with a 10 percent sodium hydroxide solution using clean paper towels. The surfaces are considered to be free of greasy contamination when water is found to spread on them evenly with a zero advancing contact angle. The water is removed by flushing the specimen with acetone. The specimens always appear shiny and metallic so that the oxide film present is not of sufficient thickness to be visible to the unaided eye. The cleaning operations are performed with the specimens mounted on their mandrels so that the mandrels can be easily inserted in the friction apparatus without touching the specimens.

The principal lubricant used in these experiments was a white mineral oil (Nujol) sold commercially for medicinal purposes. It was used without further purification, and there is some indication that it always contains a small percentage of oxidation products. The other lubricants used were of technical purity grade. The purity of these lubricants was considered sufficient for the present purpose, which was only to show the lack of correlation between their measured friction and their bulk viscosity.

In performing the friction experiments under lubricated conditions, a drop of the lubricant is applied to the specimens in the vicinity of the

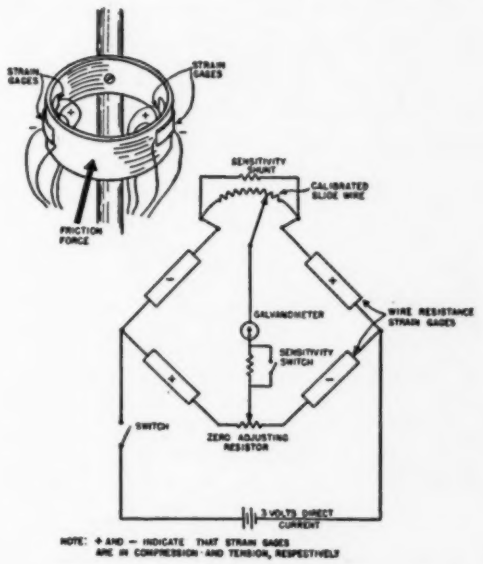


FIG. 2. Schematic diagram of friction-force measuring circuit.

point of contact. Even when the moving cylinder is rotating, surface tension keeps a sufficient quantity of the liquid surrounding the point of contact to provide an adequate excess or reservoir of lubricant.

All data presented in this paper were obtained with the stationary specimen or rider running over the same track during a given test.

A friction test consists of measuring the friction force for a series of loads at a given rubbing speed and then repeating this for a number of speeds. The general procedure consists of rotating the moving specimen at a low speed and measuring the friction forces as additional loads are applied to the specimens. Then all but the lightest loading weight is removed, the lathe is shifted to the next higher speed, and the procedure repeated. It was found that the order of speeds could be reversed or taken at random without affecting the results. Hence for simplicity the ascending order was generally followed. The positions of the specimens were not shifted, so that the rubbing areas remained the same during a single test covering both the ranges of loads and speeds. That the length of time of running on the same contact areas had no effect on the measured friction was checked by varying the orders of loads and speeds.

DEPENDENCE OF FRICTION FORCE ON LOAD

The over-all friction coefficients, obtained by dividing the total measured friction force by the total applied load at 1050 grams and 3150 grams for the whole range of speeds, are shown in Fig. 3, using white mineral oil. It is seen that these

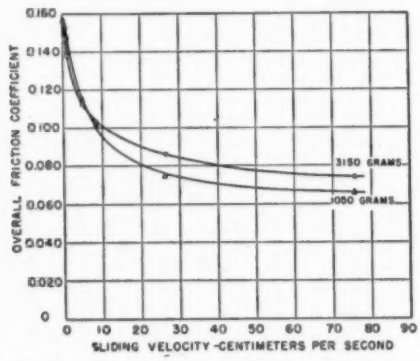


FIG. 3. Variation in the over-all friction coefficient with sliding velocity for two values of applied load. Two friction specimens of superfinished hardened steel, lubricated with white mineral oil (Nujol).

friction coefficients vary with load and decrease with increasing speed as observed by Wilson and Barnard,[4] Forester,[5] and Bristow.[6]

Examination of the region of contact on the stationary specimen after these tests disclosed a slightly worn area approximately three to five hundredths of a centimeter in diameter. A photomicrograph of a typical worn contact area is shown in Fig. 4.

If the total measured friction force is plotted against load for a given speed, curves like those in Fig. 5 are obtained. These data are seen to lie quite accurately on straight lines, except for the lightest loads where there may be some scatter, but they do not generally pass through the origin. In a single test at a given speed the loads can be taken in any order, as stated above, and can be repeated without affecting the friction force values, i.e., a single straight line is obtained. If, however, a second test is run at the same speed, using new contact areas on the same or similar specimens and a new drop of the same lubricant, then a different straight line often results, always parallel to the first, but not, in general, coincident with it. This is illustrated by the two 75.6-cm/sec. curves in Fig. 5, which were obtained by purposely varying the oil-drop size. The intercepts which the straight lines of different runs under the same conditions make with the axes do not seem to be related exactly to any of the observable conditions of the run. As regards runs at different speeds, the positive intercept on the friction force axis appeared generally to decrease with increasing speed.

This type of result has been duplicated hundreds of times with a variety of lubricants on friction specimens of different compositions and roughnesses but always having at least medium or greater hardness. To date, the only observed instances of deviation from this straight-line relationship of friction force *vs.* load have occurred when the supply of certain of the lubricants used became obviously depleted, so that the surfaces ran partially dry before the end of a test, or when the structure of the friction specimen was not homogeneous, for instance, if it comprised a thin layer or plate of hard material on top of a softer one.

The universal straight-line relation between friction force and load, illustrated by the curves

of Fig. 5, suggests using the slope of these lines as the definition of a coefficient of friction for conditions of the type described in these experiments. Let us denote it by f_Δ so that

$$f_\Delta = \Delta F/\Delta W \quad \text{for} \quad W > W_0 \qquad (1)$$

and is numerically equal to the slope of the friction force *versus* load line such as shown in Fig. 5. It can be considered to be an incremental friction coefficient and measures the rate at which the friction force increases with load once some minimum or threshold load W_0 is exceeded.

f_Δ has the obvious advantage of being independent of load and, as we shall see below, it possesses most of the characteristics usually attributed to the boundary friction coefficient. Furthermore, since it excludes the region of very small loads it should be particularly useful in any study of boundary lubrication phenomena since it eliminates the variable and inconsistent effects arising from the measurement of friction force when surfaces are brought into the lightest contact.

The term "boundary friction" is usually applied to those situations in which friction is observed to be independent of velocity (see for instance Beeck, Givens and Smith[16]). This has been generally taken as evidence that it possesses no hydrodynamic component, although such a conclusion may certainly be open to question. However, we shall undertake to show more directly that there is no hydrodynamic element present in the incremental friction coefficient by Eq. (1).

HYDRODYNAMIC EFFECTS

To begin with, the geometry of crossed cylinders was purposely chosen to minimize, as much as possible, any hydrodynamic effects. If, however, there is a hydrodynamic element present in these friction experiments, it can be expected to produce two results. The first is a viscous drag occurring in the vicinity of the contact area and within the oil drop. This drag will increase the the total friction force over that due solely to boundary action. Secondly, there should be a hydrodynamic lift due to the wedge-shaped oil film resulting from curvature of the rotating cylinder on the upstream side of the contact area. Even though it is insufficient to separate the two surfaces against the action of the applied load, it will still be present and contribute its share toward supporting the total load. Hence, each of these forces must be subtracted from the total friction force and total applied load respectively to give the non-fluid friction force arising from that part of the load that is not supported by fluid pressures. It is the ratio of these two latter quantities that gives a friction coefficient that is independent of fluid effects.

Analysis of the hydrodynamic performance of the present system shows that for rigid surfaces the hydrodynamic lift, W_f, and the viscous friction force, F_f, can depend only on the following quantities: the speed, U, the viscosity, μ, the radii of the friction cylinders, R, and the mimimum oil thickness, h_0. Hence, from dimensional analysis there must be a functional relationship between h_0/R and $W_f/\mu UR$ and between h_0/R and $F_f/\mu UR$. Therefore, if h_0/R remains constant under certain conditions, then both the quantities $W_f/\mu UR$ and $F_f/\mu UR$ must also remain constant. Now in a single test at a given speed, and using a lubricant of a given viscosity but

FIG. 4. Photomicrograph of a typical worn contact area. Superfinished hardened steel specimens lubricated with white mineral oil (Nujol). Running time of 20 minutes at 400-g. total applied load and 75.7-cm./sec. sliding velocity. 100×.

[15] J. Howlett, J. App. Phys. **17**, 137 (1946).
[16] O. Beeck, J. W. Givens, and A. E. Smith, Proc. Roy. Soc. **A177**, 90 (1940).

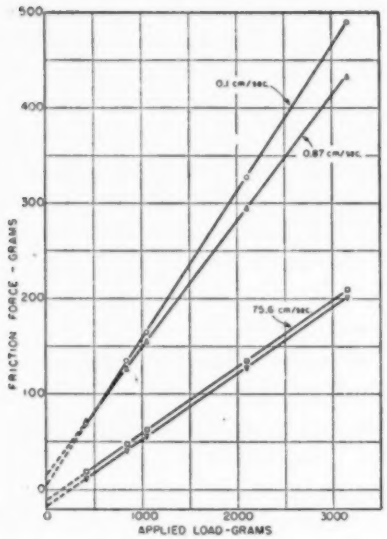


FIG. 5. Variation in measured friction force with applied load for several sliding velocities. Superfinished hardened steel specimens lubricated with white mineral oil (Nujol).

with varying loads, the appearance of the worn contact areas, even at the lightest loads, gives definite evidence that the surfaces are always in contact above a certain minimum load and, hence, that h_0 does not change. This is shown in Fig. 4 for a twenty-minute run at a load of 400 grams, using a mineral oil lubricant, and the highest speed of 75.7 cm/sec. Therefore, in a single test one would expect the hydrodynamic contributions, F_f and W_f, to the total friction force and load carrying capacity to be constant for all applied loads. Hence, they cannot affect the slope of the lines of Fig. 5 which define the incremental friction coefficient. (This argument will not be entirely valid if one takes account of local elastic deformation of the surfaces in the vicintiy of the area of contact, but little can be surmised about its effect except that it is probably small.)

It is possible to make a quantitative estimate of the size of these hydrodynamic effects. While the hydrodynamic performance of a system with this geometry has not been analyzed mathematically, Howlett[15] has analyzed a related problem, namely, that of a rotating sphere in contact with a stationary one. If the axis of rotation is at right angles to the line joining the centers of the two spheres, then he obtains the following relation between the minimum film thickness, h_0, and the applied load, W:

$$h_0 = 36/25 \cdot r^5(\mu\omega/W)^2, \qquad (2)$$

where r is the radius of both spheres and ω the angular speed of rotation. While this geometry is not the same as two crossed cylinders, it can be expected to be similar, especially in the vicinity of the contact areas. If we take r as equal to twice the radius of the cylinders, this should make the wedge angle of the two oil films equivalent. The viscosity of the white mineral oil used in these experiments is 1.7 poises at room temperature and ω is 69 radians/sec., corresponding to the speed of 75.7 cm/sec. Using these values in Eq. (2), we obtain $h_0 = 1.0 \times 10^{-6}$ centimeter at a load of 1000 grams and 1.0×10^{-4} centimeter at a load of 100 grams. Since the former figure is less than the height of the surface irregularities, a bulk fluid film of this thickness probably has no significance. Hence part of the load, at least, must be carried by boundary lubricated areas. On the other hand, a film thickness of 1.0×10^{-4} cm or 40 micro-inches probably is significant and indicates complete separation of the surfaces. This is in line with the observed deviation of the friction force values from a straight line at loads below 400 grams.

From the above argument, one may conclude that the fluid effects are responsible, at least in part, for the intercept of the friction force curves shown in Fig. 5. This explains why the intercept varies from run to run under the same conditions of load and speed and is probably due to slight differences in the size of the oil drop.

Further evidence that the fluid effects do not vary with load for a given speed is the previously mentioned observation that the friction force does not change with the size of the wear spot, i.e., with the length of time of rubbing. This suggests that a change in the fluid film wedge angle at the edge of the contact area as a result of wear does not appreciably affect the fluid drag or lift. It further indicates that any fluid effects which may exist *within* the contact area, as has been recently suggested by several writers,* do not vary with the size of the apparent contact area over a range of a 100-fold in the latter. This

* See, for instance, Proceedings of M.I.T. Summer Conference on Mechanical Wear, June 1948.

suggests that such fluid effects, if they exist, cannot contribute much to the total friction.

Having established that the friction coefficient as defined by Eq. (1) is devoid of any hydrodynamic element, it may be of interest to see how it varies with speed. In Fig. 6 this incremental friction coefficient between similar steel surfaces is plotted against speed for a number of different lubricants. These data are replotted in Figs. 7(a) and 7(b) on a semilogarithmic scale. It is interesting to see that it varies with speed over at least part of the range, much as did the over-all friction coefficient plotted in Fig. 3, decreasing with increasing speed.

In spite of the above discussion, this decreasing trend still suggests a hydrodynamic effect, the higher speed seemingly introducing an increasing proportion of fluid friction with its correspondingly smaller coefficient. It is, therefore, of interest to compare the curves in Fig. 7(b) for butyl stearate and the two glycerin-water mixtures. The viscosity of the butyl stearate at room temperature of 24°C. was found, by direct measurement in a Saybolt viscosimeter, to be 8.0 centipoise. A glycerin-water mixture was prepared having a viscosity of 8.0 centipoise, as determined in the same viscosimeter at 24°C. The composition of the mixture was approximately 58 percent glycerin by weight. A second glycerin-water mixture, containing approximately 95 percent glycerin by weight, was prepared having a viscosity of 394 centipoise at 24°C. It can be seen from Fig. 7(b) that the curves of incremental friction coefficient *versus* speed for the two glycerin-water mixtures are identical within the experimental error of the method, even though their viscosities differ by a factor of 50. On the other hand, the curves obtained with two fluids of identical viscosities, butyl stearate and the 58 percent glycerin-water mixture, are quite different in both value and shape. Furthermore, in spite of the very low value for friction coefficient (less than 0.02) of the 95 percent glycerin-water mixture at the high speed, a photomicrograph of the contact area for a twenty-minute run at this speed and the lightest load of 400 grams, reproduced in Fig. 8, shows the typical signs of a metal-to-metal contact during running. Hence, this complete absence of correlation between the friction coefficient curves

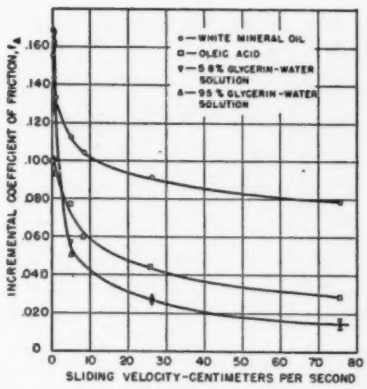


FIG. 6. Variation in the incremental coefficient of friction with sliding velocity for several lubricants. Superfinished hardened steel specimens.

and the viscosities of the lubricants, together with the evidence of solid contact in the case of the most viscous fluid, indicates that the incremental friction coefficient is completely independent of the effects of fluid film lubrication. Its decrease with speed, therefore, is an inherent characteristic of non-fluid lubricated friction between surfaces of this type.

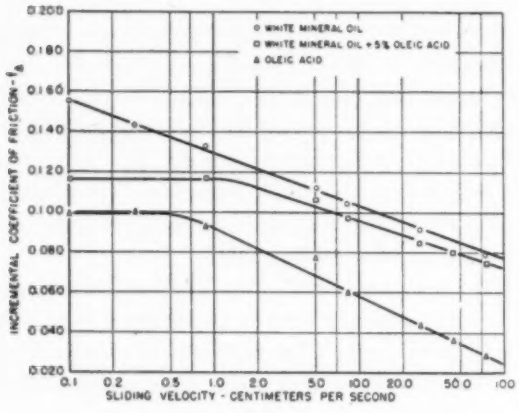
To make sure that this decrease was not an inherent characteristic of the apparatus, such as might be caused by the stationary specimen being thrown out of contact with the moving one part of the time at the high speeds, some tests were run using similar unlubricated specimens. In this case there was considerable vibration as a result of severe galling, but the friction coefficient, for an applied load of 630 grams, was found to remain constant at 0.93 over the entire range of speed.

DEPENDENCE OF FRICTION COEFFICIENT ON SPEED

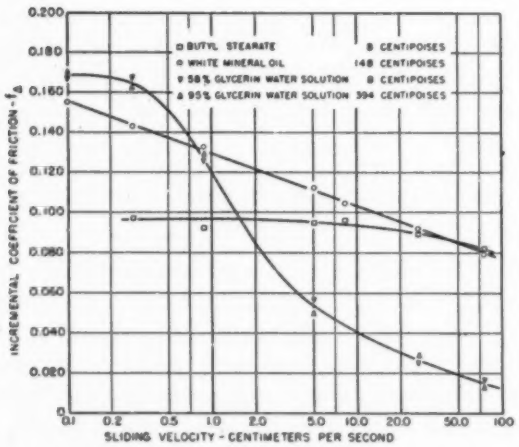
So far the discussion has been principally concerned with whether hydrodynamic effects could be present in the curves of Figs. 6 and 7. The curves themselves, however, are of considerable interest. The salient characteristics are seen most clearly on the semilogarithmic scale of Fig. 7. All of the curves decrease with increasing speed in the upper speed range; some decrease over the whole speed range, while others, like oleic acid, are constant at the lower speeds.

Beeck, Givens, and Smith[16] have obtained, in a different way, curves somewhat similar to that

shown in Fig. 7 for oleic acid. They define the flat portion where the friction is independent of the speed as the "boundary" region of operation, and they define the sloping part of the curves as a region of "quasi-hydrodynamic" operation because of the variation of friction with speed. They did not elucidate the mechanism in this latter region very much except to suggest that it is due to a wedging action of the molecules produced by oil drag. They found, generally, that lubricants which exhibited this wedging effect showed preferred orientation on a solid surface, as evidenced by electron diffraction, and that the effect was independent of the bulk viscosity. They further found that a high degree of polish of the surfaces was necessary for the effect.


WHITE MINERAL OIL
WHITE MINERAL OIL + 5% OLEIC ACID
OLEIC ACID
INCREMENTAL COEFFICIENT OF FRICTION - f_Δ
SLIDING VELOCITY - CENTIMETERS PER SECOND

(a)


BUTYL STEARATE 8 CENTIPOISES
WHITE MINERAL OIL 148 CENTIPOISES
58% GLYCERIN WATER SOLUTION 8 CENTIPOISES
95% GLYCERIN WATER SOLUTION 394 CENTIPOISES
INCREMENTAL COEFFICIENT OF FRICTION - f_Δ
SLIDING VELOCITY - CENTIMETERS PER SECOND

(b)

FIG. 7. Semilogarithmic plot of variation in incremental coefficient of friction with sliding velocity for several lubricants. Superfinished hardened steel specimens. (a) Lubricated with mineral oil (Nujol), mineral oil plus 5 percent oleic acid, and pure oleic acid. (b) Lubricated with 95 percent glycerin-water mixture, 58 percent glycerin-water mixture, white mineral oil (Nujol), and butyl stearate.

There are, however, some important experimental differences in the characteristics of the curves of Beeck, Givens, and Smith and those of Figs. 6 and 7. First, the velocity at which the friction coefficient ceased to be constant was found by these authors to be somewhat evanescent. Their method consisted of measuring the deceleration of a four-ball machine, and they found that this "critical velocity" did not appear if the initial speed of the decelerating top was much higher than the critical, and that it could never be observed on accelerating. On the other hand, the breaks in the oleic acid and butyl stearate curves of Fig. 7 were quite reproducible regardless of the order in which the speeds were taken. Finally, although these authors' curves show that with white mineral oil the friction force decreases with increasing speed at low speeds, much as in Fig. 7, yet they state that this oil exhibits a critical velocity at a comparatively high speed. Hence, although the observations of of Beeck, Givens, and Smith, and those of the present study may be describing somewhat similar phenomena, it seems inappropriate to apply the term "quasi-hydrodynamic" to the present variation of friction with speed, since the worn contact areas suggest that the rubbing surfaces remain in contact even at the highest speeds employed.

It is not possible at the present time to offer very specific explanation of the decrease of f_Δ with the speed, but the following tentative one is suggested. It seems to be a fact that the friction of all the polar lubricants so far examined shows an independence of speed at the lower speeds, whereas that of the non-polar lubricants generally does not do so. Furthermore, the constant values of friction coefficient of the polar compounds at low speeds is always lower than that of the non-polar compounds at the same speeds. However, the latter can achieve as low a value of friction coefficient as the constant values of the former at a sufficiently higher speed. This suggests that the effect of speed is to impart to the non-polar lubricant film at higher speeds that same property which polar lubricants already possess at lower speeds, namely, that of orienta-

tion and close packing of the molecules in the monolayer adjacent to the surface.** In addition, if the non-polar molecules do not possess any orientation at the lowest speed and speed induces orientation, then such molecules should show a dependence of friction on speed at all speeds. On the other hand, speed will only increase the degree or amount of orientation of the polar molecules above a certain minimum speed, since below that speed the degree of orientation which speed would have induced has already been achieved by the electric field of the solid surface acting on the polar molecules. Beeck, Givens, and Smith found that the straight-chain hydrocarbon, tetratriacontane ($C_{34}H_{70}$), exhibited the wedging effect, and that a well rubbed film of it gave an electron diffraction pattern showing preferred orientation. Brummage[17] likewise found by electron diffraction that rubbed films of the normal paraffins tetracosane ($C_{24}H_{50}$) and triacontane ($C_{30}H_{62}$), as well as tetratriacontane, all showed preferred orientation of the molecules. Therefore, one might expect that the smaller hydrocarbon molecules, such as are present in white mineral oil, would take up preferred orientation under the action of rubbing, although being more mobile this orientation would not persist and therefore could not be observed in a diffraction pattern after the rubbing had ceased.

A second possible explanation of the decrease of f_Δ with speed is as follows: Bowden, Moore, and Tabor[18] suggest that in contrast to dry friction the action of a lubricant is to interpose on the two surfaces films whose mutual shear strength is less than that of the clean metal. They have evidence, however, to show that this substitution is never complete, so that a certain percentage of the total friction force is still caused by local adhesions between the bare metals. The efficacy of the lubricant depends in part on its ability to reduce the number and area of these metallic junctions. This suggests that the effect of speed on f_Δ may be to increase the fraction of the local load-carrying contact areas that are kept coated with the lubricant. Since, however, such coating efficiency is generally thought to depend to some extent on the formation of close-packed monolayers, this hypothesis may reduce to simply another aspect of the one first presented.

FIG. 8. Photomicrograph of the worn contact area obtained during a twenty-minute run using 95 percent glycerin-water mixture as the lubricant. Total applied load of 400 grams and sliding velocity of 75.7 cm/sec. Specimens of superfinished hardened steel. 100×.

Still a third possible explanation of the decrease of f_Δ with speed is that the resulting shear breaks down a long-range ordering or orientation many molecular layers thick induced by the solid surfaces. The existence of such long-range orientation in the vicinity of solid surfaces is by no no means completely accepted, but the sum of evidence for it seems to be increasing.*** One trouble with its application to the present observations is that the resultant rigidity and shear resistance would be expected to be greatest with the polar molecules such as oleic acid and butyl

** A similar suggestion has recently been made independently by Larsen and Perry in a paper presented at the 1948 M.I.T. Summer Conference on Mechanical Wear.

[17] K. G. Brummage, Proc. Roy. Soc. **A188**, 414 (1947).

[18] F. P. Bowden, A. J. W. Moore, and D. Tabor, J. App. Phys. **14**, 80 (1943).

*** See, for instance the, recent critical survey of the subject by J. C. Henniker and J. W. McBain conducted for the Office of Naval Research (*Properties of Surfaces:* N6 ori-154-T.O. II, March 31, 1948.)

stearate, whereas at the lowest speeds Fig. 7 shows that the reverse is true.

Much more investigation needs to be made of the dependence of this friction coefficient on speed for a wider variety of lubricants and surfaces before a more complete picture can be given. This work is being continued, particularly on the effect of composition, hardness, and roughness of the surfaces, as well as the effect of the lubricant. However, from the results so far obtained it is believed that the rather striking characteristics of the curves of Fig. 7 are exhibited most clearly when the friction surfaces are quite hard and smooth, as they are in the present experiments. Plastic deformation of soft or rough surfaces may be expected to obscure these characteristics which Fig. 7 shows are definitely properties of the lubricant.

It is also interesting to note in Fig. 7 that the decrease of friction coefficient when it occurs seems to be generally proportional to the logarithm of the speed. There is no explanation for this at present.

A word might be said about the curves for the glycerin-water mixtures of Fig. 7. They were chosen originally to test the effect of viscosity. Their low value of f_Δ, particularly at the high speeds, is quite surprising, as glycerin is generally not considered to be a good boundary lubricant. The values at the lowest speeds are not inconsistent with a value of 0.2 for the static value given by Wilson and Barnard.[4] Leloup[19] has obtained similar curves for glycerin, but as he used a sleeve bearing and measured the total friction coefficient, the friction was undoubtedly largely hydrodynamic at the higher speeds. At the present time there is no explanation for the low values of its friction at the high speeds.

In concluding, it should be noted that we have avoided calling f_Δ, as defined by Eq. (1), a boundary friction coefficient. "Boundary" friction is defined in various ways by various workers. This incremental friction coefficient has most of the characteristics of a boundary friction coefficient, if not all of them. For instance, it has been shown to be independent of fluid effects. However, it is defined in a rather special way, namely, as the slope of the linear curve of friction force *versus* applied load, and should not be identified completely with a boundary friction coefficient until we know more completely what the latter is. For instance, if the conventional boundary friction force contains a component that is independent of load, this component of the force will be excluded from f_Δ because of the way in which f_Δ is determined.

Finally, it should also be emphasized that in no sense is f_Δ the *complete* friction coefficient, that is, it is not the ratio of the *total* measured friction force to the *total* applied load, even under the conditions of these experiments, although for the higher loads it approximates the over-all friction coefficient. It has been defined and selected for study in the hope that it will prove useful in separating and revealing the roles of lubricant and of bearing material in non-fluid lubricated friction.

[19] L. Leloup, Rev. Univ. d. Mines **3**, 373 (1947), ninth series.

SUMMARY AND ACKNOWLEDGMENT

To summarize, the incremental friction coefficient, defined by Eq. (1), is found to be independent of hydrodynamic effects for the following reasons:

(a) For a given speed the separation of the surfaces remains unchanged over the range of applied loads; in fact, the surfaces remain always in contact. The evidence for this is the occurrence of visible wear spots at all loads. If the separation does not vary with load then the hydrodynamic effects must also remain constant, and because of this their effects are canceled out in the calculation of f_Δ.

(b) The independence of the friction force, measured at one speed and load, of the size of the contact area suggests that any fluid effects occurring *within* the contact area contribute little to the total friction, and hence may be neglected.

(c) There is a complete lack of correlation between the viscosity of the lubricant and the curve of incremental friction coefficient *versus* sliding velocity.

In spite of the absence of hydrodynamic effects, the incremental friction coefficient decreases with increasing speed over part or all of the range of speed for various lubricants. The general character of this relation seems to be

characteristic of the particular lubricant. The rate of decrease when it occurs is roughly proportional to the logarithm of the speed.

The incremental friction coefficient does not include all the components of friction force present under the usual conditions of boundary operation. However, this very fact should render it very useful as an investigative tool, since it does not represent the resultant of a large number of diverse effects.

This work is made possible through the generosity of the Chrysler Corporation.

The Optical Equations of Three-Dimensional Photoelasticity

RAYMOND D. MINDLIN AND LAWRENCE E. GOODMAN
Department of Civil Engineering, Columbia University, New York, New York
(Received July 12, 1948)

The fundamental optical problem encountered in three-dimensional photoelasticity is that of the propagation of light in a heterogeneous, aeolotropic medium. Although there is an extensive literature on the homogeneous, aeolotropic medium and the heterogeneous, isotropic medium, only a few special cases of the combined problem have been studied from the point of view of Maxwell's electromagnetic equations. With regard to approximate treatments, the most general is that of F. Neumann (1841), derived from purely kinematical considerations. In the present paper the passage from Maxwell's to Neumann's equations is executed and the nature and order of magnitude of the necessary approximations are disclosed. During the passage, there are discovered a pair of equations, of relatively simple form, which open the way to a variety of new solutions in closed form and reveal the essential parameter upon which the optical phenomena depend.

I. INTRODUCTION

IN applying the photoelastic method of stress analysis to states of stress which vary in three dimensions, the experimenter is confronted with the problem of interpreting the effects observed when light is passed through a "crystal" whose properties vary from point to point in a very general manner. The dynamical equations governing the propagation of light in homogeneous-aeolotropic media and in heterogeneous-isotropic media have been explored extensively but there appear to be only two exact solutions of the equations for cases in which the medium is both heterogeneous and aeolotropic. One of these[1] is for the case in which the principal axes of the section of the index ellipsoid by a plane perpendicular to the wave normal vary linearly along the wave normal. Another[2] is for the case in which the principal axes of the section of the index ellipsoid rotate as a linear function of the distance along the wave normal. Both of these solutions differ little from the analogous solution of the approximate equations whose derivation is based upon simple kinematical considerations.

As more complex states of stress are contemplated, solutions of the dynamical equations become increasingly difficult and it is very likely that, even if they were found, they would differ only slightly from the corresponding solutions of the kinematical equations. However, since the two types of equations are derived quite independently, the nature and order of magnitude of the approximations involved in the kinematical equations are not apparent at the outset. One purpose of this paper is to reveal these approximations by executing the passage from the dynamical to the kinematical equations. This is accomplished in two steps. First it is shown that an approximation, usually ignored, is involved in the type of medium produced by a two-dimensional state of stress. This approximation is adopted at the outset of the three-dimensional considerations and the additional approximations are then disclosed.

II. ELECTROMAGNETIC AND STRESS-OPTICAL EQUATIONS

For the dynamical system we choose Maxwell's electromagnetic equations. These become, for

[1] R. D. Mindlin, J. App. Mech. **8**, A187 (1941).
[2] D. C. Drucker and R. D. Mindlin, J. App. Phys. **11**, 724 (1940).

our case,[3]

$$\dot{\mathbf{D}}=\nabla\times\mathbf{H}, \qquad \nabla\cdot\mathbf{H}=0, \tag{II.1}$$

$$\dot{\mathbf{B}}=-\nabla\times\mathbf{E}, \qquad \nabla\cdot\mathbf{B}=0, \tag{II.2}$$

$$\mathbf{B}=\mu\mathbf{H}, \tag{II.3}$$

$$\mathbf{E}=\mathbf{\Pi}\cdot\mathbf{D}, \tag{II.4}$$

where $\mathbf{E}$ is the electric vector, $\mathbf{H}$ is the magnetic vector, $\mathbf{D}$ is the electric induction, $\mathbf{B}$ is the magnetic induction and μ is the (constant) magnetic permeability. In the electrostatic system of units $\mu=v_0^{-2}$, where v_0 is the velocity of light in vacuum. The heterogeneity and aeolotropy of the medium are characterized by the self-conjugate index dyadic $\mathbf{\Pi}$ whose elements are the reciprocals of the squares of the indices of refraction n_{rs} $(r, s=1, 2, 3)$. Thus, if $\mathbf{r}$ is the position vector, the index (Fresnel) ellipsoid is given by the equation

$$\mathbf{r}\cdot\mathbf{\Pi}\cdot\mathbf{r}=1. \tag{II.5}$$

In a medium possessing isotropic elasticity, the stress-optical relation may be written as

$$\mathbf{\Pi}-n_0^{-2}\mathbf{I}=C\mathbf{\Psi}+C'\Theta\mathbf{I}, \tag{II.6}$$

where $\mathbf{I}$ is the idemfactor, C and C' are stress-optical coefficients, $\mathbf{\Psi}$ is the stress dyadic and Θ its first invariant. The symbol n_0 represents, at the least, the index of refraction of the unstressed material, but may include a symmetric function of the principal stresses. Equation (II.6) is equivalent to the law proposed by Coker and Filon.[4]

It is convenient, for our purposes, to write the equations of motion in terms of the magnetic vector. From (II.2) and (II.3),

$$\mu\dot{\mathbf{H}}=-\nabla\times\mathbf{E}. \tag{II.7}$$

From (II.7) and (II.4)

$$\mu\dot{\mathbf{H}}=-\nabla\times(\mathbf{\Pi}\cdot\mathbf{D}). \tag{II.8}$$

Finally, operating on (II.8) with $\partial/\partial t$ and using (II.1),

$$\mu\ddot{\mathbf{H}}=-\nabla\times(\mathbf{\Pi}\cdot\nabla\times\mathbf{H}), \quad \nabla\cdot\mathbf{H}=0. \tag{II.9}$$

Equations (II.9), with $\mathbf{\Pi}$ given by (II.6), are the dynamical equations governing the propagation of light in our medium.

III. TWO-DIMENSIONAL CONSIDERATIONS

Let us suppose, first, that the stress dyadic is a function of x and y only and has one principal axis everywhere parallel to z. Identifying the axes 1, 2, 3 of the index dyadic with the rectangular coordinate axes x, y, z respectively, we have, from (II.6),

$$n_{13}^{-1}=n_{31}^{-1}=n_{23}^{-1}=n_{32}^{-1}=0, \tag{III.1}$$

$$n_{rs}=n_{rs}(x, y). \tag{III.2}$$

Using (III.1) and (III.2), Eqs. (II.9) become the following four equations governing u, v, w (the x, y, z components of $\mathbf{H}$):

$$\mu\frac{\partial^2 u}{\partial t^2}=n_{22}^{-2}\frac{\partial^2 u}{\partial z^2}-n_{21}^{-2}\frac{\partial^2 v}{\partial z^2}-n_{22}^{-2}\frac{\partial^2 w}{\partial x\partial z}+n_{21}^{-2}\frac{\partial^2 w}{\partial y\partial z}-\frac{\partial}{\partial y}\left[n_{33}^{-2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)\right], \tag{III.3}$$

$$\mu\frac{\partial^2 v}{\partial t^2}=n_{11}^{-2}\frac{\partial^2 v}{\partial z^2}-n_{12}^{-2}\frac{\partial^2 u}{\partial z^2}-n_{11}^{-2}\frac{\partial^2 w}{\partial y\partial z}+n_{12}^{-2}\frac{\partial^2 w}{\partial x\partial z}+\frac{\partial}{\partial x}\left[n_{33}^{-2}\left(\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}\right)\right], \tag{III.4}$$

$$\mu\frac{\partial^2 w}{\partial t^2}=-\frac{\partial}{\partial x}\left[n_{21}^{-2}\left(\frac{\partial w}{\partial y}-\frac{\partial v}{\partial z}\right)+n_{22}^{-2}\left(\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x}\right)\right]+\frac{\partial}{\partial y}\left[n_{11}^{-2}\left(\frac{\partial w}{\partial y}-\frac{\partial v}{\partial z}\right)+n_{12}^{-2}\left(\frac{\partial u}{\partial z}-\frac{\partial w}{\partial x}\right)\right], \tag{III.5}$$

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}=0. \tag{III.6}$$

Anticipating solutions in the form of waves travelling in the z direction it is awkward to have the derivatives, with respect to z, of both u and v in each of (III.3) and (III.4). This complication may be eliminated by using the stress trajectories as a system of orthogonal curvilinear coordinates.

The central section of the ellipsoid (II.5) by

[3] L. N. G. Coker and E. G. Filon, *A Treatise on Photo-Elasticity* (Cambridge University Press, Teddington, 1931), Chap. I.

[4] See reference 3, p. 198, Eq. (3.102).

a plane $z=$const. is the ellipse

$$n_{11}^{-2}x^2+2n_{12}^{-2}xy+n_{22}^{-2}y^2=1. \quad \text{(III.7)}$$

The principal semi-axes of (III.7) have lengths given by the stationary values of

$$n_N^{-2}=n_{11}^{-2}\cos^2\theta+n_{22}^{-2}\sin^2\theta +2n_{12}^{-2}\sin\theta\cos\theta, \quad \text{(III.8)}$$

where θ is the angle between the x-axis and a direction N in the x, y plane. The stationary values occur at $\theta=\varphi$, $\varphi+\pi/2$, where

$$\tan 2\varphi=2n_{12}^{-2}/(n_{11}^{-2}-n_{22}^{-2}), \quad \text{(III.9)}$$

and have magnitudes

$$n_a^{-2}=[n_N^{-2}]_{\theta=\varphi}=n_{11}^{-2}\cos^2\varphi+n_{22}^{-2}\sin^2\varphi +n_{12}^{-2}\sin 2\varphi, \quad \text{(III.10)}$$

$$n_b^{-2}=[n_N^{-2}]_{\theta=\varphi+\pi/2}=n_{11}^{-2}\sin^2\varphi +n_{22}^{-2}\cos^2\varphi-n_{12}^{-2}\sin 2\varphi, \quad \text{(III.11)}$$

where n_a and n_b are principal indices of refraction, in view of the present restrictions on Ψ. The principal velocities a, b, c are given by

$$a^2=\mu^{-1}n_a^{-2}, \quad b^2=\mu^{-1}n_b^{-2}, \quad c^2=\mu^{-1}n_{33}^{-2}. \quad \text{(III.12)}$$

Our curvilinear coordinates s_1, s_2 increase along curves whose tangents make angles φ and $\varphi+\pi/2$, respectively, with the x axis. Hence

$$\frac{\partial}{\partial x}=\cos\varphi\frac{\partial}{\partial s_1}-\sin\varphi\frac{\partial}{\partial s_2}, \qquad \frac{\partial}{\partial y}=\sin\varphi\frac{\partial}{\partial s_1}+\cos\varphi\frac{\partial}{\partial s_2}. \quad \text{(III.13)}$$

Finally, let u' and v' be the components of $\mathbf{H}$ in the directions s_1 and s_2 (increasing) respectively. Then

$$u=u'\cos\varphi-v'\sin\varphi, \; v=v'\cos\varphi+u'\sin\varphi. \quad \text{(III.14)}$$

Using (III.9) to (III.14), Eqs. (III.3) and (III.4) become

$$\frac{\partial^2 u'}{\partial t^2}-b^2\frac{\partial^2 u'}{\partial z^2}=b^2\frac{\partial\Delta_1}{\partial s_1}-\frac{\partial(c^2\omega_1)}{\partial s_2}, \quad \text{(III.15)}$$

$$\frac{\partial^2 v'}{\partial t^2}-a^2\frac{\partial^2 v'}{\partial z^2}=a^2\frac{\partial\Delta_1}{\partial s_2}+\frac{\partial(c^2\omega_1)}{\partial s_1}, \quad \text{(III.16)}$$

where

$$\Delta_1=-\frac{\partial w}{\partial z}=\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y} =\frac{\partial u'}{\partial s_1}+\frac{\partial v'}{\partial s_2}+u'\frac{\partial\varphi}{\partial s_2}-v'\frac{\partial\varphi}{\partial s_1}, \quad \text{(III.17)}$$

$$\omega_1=\frac{\partial v}{\partial x}-\frac{\partial u}{\partial y}=\frac{\partial v'}{\partial s_1}-\frac{\partial u'}{\partial s_2}+v'\frac{\partial\varphi}{\partial s_2}+u'\frac{\partial\varphi}{\partial s_1}. \quad \text{(III.18)}$$

It is customary, in two-dimensional photoelasticity, to employ solutions of the type

$$u'=A\cos(\psi-\varphi)\cos\omega[t-(z/b)], \quad \text{(III.19)}$$

$$v'=A\sin(\psi-\varphi)\cos\omega[t-(z/a)], \quad \text{(III.20)}$$

where ω is the frequency and A is proportional to the intensity of the incident light and ψ is the (constant) angle between the x axis and the direction of polarization at incidence. It may be seen that (III.19) and (III.20) are solutions of (III.15) and (III.16) only for homogeneous stress (which transforms the material into a homogeneous crystal) whereupon the right-hand sides of (III.15) and (III.16) vanish.

To find the nature and order of magnitude of the terms that are ordinarily neglected, substitute (III.19) and (III.20) in (III.15) and (III.16) and consider the ratios of the amplitudes of the terms on the right to those on the left. These ratios are of the following six types:

$$\text{(i)} \quad \frac{b^2}{\omega^2}\frac{\partial^2\varphi}{\partial s^2}=O\left[\lambda^2\frac{\partial^2\varphi}{\partial s^2}\right],$$

$$\text{(ii)} \quad \left(\frac{b}{\omega}\frac{\partial\varphi}{\partial s}\right)^2=O\left[\lambda\frac{\partial\varphi}{\partial s}\right]^2,$$

$$\text{(iii)} \quad \left(\frac{b}{\omega}\frac{\partial\varphi}{\partial s}\right)\left(bz\frac{\partial b^{-1}}{\partial s}\right)=O\left[\lambda z\frac{\partial\varphi}{\partial s}\frac{\partial n}{\partial s}\right],$$

$$\text{(iv)} \quad \frac{b}{\omega}bz\frac{\partial^2 b^{-1}}{\partial s^2}=O\left[\lambda z\frac{\partial^2 n}{\partial s^2}\right],$$

$$\text{(v)} \quad \frac{1}{\omega^2}\frac{\partial\varphi}{\partial s}\frac{\partial c^2}{\partial s}=O\left[\lambda^2\frac{\partial n}{\partial s}\frac{\partial\varphi}{\partial s}\right],$$

$$\text{(vi)} \quad \frac{z}{\omega}\frac{\partial c^2}{\partial s}\frac{\partial b^{-1}}{\partial s}=O\left[\lambda z\left(\frac{\partial n}{\partial s}\right)^2\right],$$

where s, n and b are representative lengths in the x, y plane, index of refraction, and wave velocity, respectively, λ is the wave-length of light in vacuum and $O[\ ]$ stands for "order of magnitude of." To estimate the order of magnitude of these ratios, consider a wave-length of 6×10^{-4} mm, a stress-gradient of 10^{8} dynes/cm^{2} ($\cong$1500 lb./in.^{2}) per mm and a stress-optical coefficient of 100 Brewsters. Under these extreme conditions, $\lambda\partial n/\partial s=O[10^{-7}]$ and, for a model 5 mm thick, $z\partial n/\partial s=O[10^{-3}]$. With regard to terms involving φ, if the rotation of the principal axes were 2π radians in one millimeter, $\lambda\partial\varphi/\partial s=O[10^{-3}]$. Since the ratios involve squares or products of terms not exceeding $O[10^{-3}]$ the ratios themselves do not exceed $O[10^{-6}]$.

It is true that, in the neighborhood of an isotropic point, $\partial\varphi/\partial s$ becomes very large. However, $\partial\varphi/\partial s$ may be expressed in terms of $\partial n/\partial s$ and the principal-stress difference through the use of the stress equations of equilibrium and the stress-optical law, following which it may be shown, by a limiting process, that the ratios involving $\partial\varphi/\partial s$ are, again, small.

Hence, all of the terms on the right-hand sides of (III.15) and (III.16) are negligible and these equations may be replaced, to a good approximation, by

$$b^{2}\frac{\partial^{2}u'}{\partial z^{2}}=\frac{\partial^{2}u'}{\partial t^{2}}, \tag{III.21}$$

$$a^{2}\frac{\partial^{2}v'}{\partial z^{2}}=\frac{\partial^{2}v'}{\partial t^{2}}. \tag{III.22}$$

Equations (III.19) and (III.20) are solutions of (III.21) and (III.22) and lead directly to the usual laws of photoelasticity, namely, the directions of polarization coincide with the directions of principal stress in the plane perpendicular to the wave normal and the relative phase retardation is

$$\delta=\omega z\left(\frac{1}{b}-\frac{1}{a}\right)=\frac{2\pi z}{\lambda}(n_{b}-n_{a}). \tag{III.23}$$

It is useful to observe that the order of Eqs. (III.21) and (III.22) may be reduced if we are interested only in waves moving in one direction. In this case we may write

$$\frac{\partial u'}{\partial z}=\mp\frac{1}{b}\frac{\partial u'}{\partial t}, \tag{III.24}$$

$$\frac{\partial v'}{\partial z}=\mp\frac{1}{a}\frac{\partial v'}{\partial t}. \tag{III.25}$$

The minus signs apply to forward moving waves, such as (III.19) and (III.20) while the plus signs refer to backward moving waves which we may ignore for most purposes. It may be verified, by substitution, that (III.24) and (III.25) are indeed solutions of (III.21) and (III.22).

The restriction of z independence placed on the stress dyadic at the beginning of this section is a rather severe one if we remember that this is possible in relatively few states of stress even in "two-dimensional" photoelasticity. However, the purpose of this section is to dispose of x and y dependence so that the more difficult case of z dependence may be treated separately.

IV. THE OPTICAL WAVE EQUATIONS OF THREE-DIMENSIONAL PHOTOELASTICITY

In general the index dyadic is a function of all three coordinates. However, if we are interested in plane waves with wave normal in, say, the z direction, the results of the preceding section show that, for variations of the index dyadic that might be encountered in photoelasticity, the dependence on x and y may be neglected. In fact, a detailed study of the general equations shows that all of the x and y dependent terms are of the order of magnitude of those already considered in the preceding section. Accordingly we may assume that the n_{rs} are functions of z only and, hence, u, v and w are functions of z and t only. If these restrictions are introduced into (II.9) the differential equations reduce to $w=0$ and

$$\mu\frac{\partial^{2}u}{\partial t^{2}}=n_{22}{}^{-2}\frac{\partial^{2}u}{\partial z^{2}}-n_{21}{}^{-2}\frac{\partial^{2}v}{\partial z^{2}}+\frac{dn_{22}{}^{-2}}{dz}\frac{\partial u}{\partial z}-\frac{dn_{21}{}^{-2}}{dz}\frac{\partial v}{\partial z}, \tag{IV.1}$$

$$\mu\frac{\partial^{2}v}{\partial t^{2}}=n_{11}{}^{-2}\frac{\partial^{2}v}{\partial z^{2}}-n_{12}{}^{-2}\frac{\partial^{2}u}{\partial z^{2}}+\frac{dn_{11}{}^{-2}}{dz}\frac{\partial v}{\partial z}-\frac{dn_{12}{}^{-2}}{dz}\frac{\partial u}{\partial z}. \tag{IV.2}$$

The equation of the central section of the index ellipsoid by a plane perpendicular to the z axis is again given by (III.7). The principal axes of this section are oriented at angles φ and $\varphi+\pi/2$ with respect to the x axis, where φ is given by (III.9). It is to be noted that these axes are not in general principal axes of the index ellipsoid, as they were in Section III, but are principal axes of the central section perpendicular to the z axis. These have been called secondary principal axes. The lengths of the secondary principal semi-axes of the section are given by the stationary values, say n_1 and n_2 of (III.8) where

$$n_1^{-2}=[n_N^{-2}]_{\theta=\varphi}=n_{11}^{-2}\cos^2\varphi+n_{22}^{-2}\sin^2\varphi+n_{12}^{-2}\sin 2\varphi, \quad \text{(IV.3)}$$

$$n_2^{-2}=[n_N^{-2}]_{\theta=\varphi+\pi/2}=n_{11}^{-2}\sin^2\varphi+n_{22}^{-2}\cos^2\varphi-n_{12}^{-2}\sin 2\varphi. \quad \text{(IV.4)}$$

The velocities corresponding to n_1 and n_2 will be designated by k_1 and k_2, where

$$k_1^2=\mu^{-1}n_1^{-2}, \quad k_2^2=\mu^{-1}n_2^{-2}. \quad \text{(IV.5)}$$

It is convenient to refer (IV.1) and (IV.2) to secondary principal axes. Using (IV.3) to (IV.5) and (III.14), we find

$$\frac{\partial^2 u'}{\partial t^2}=\frac{\partial}{\partial z}\left[k_2^2\left(\frac{\partial u'}{\partial z}-v'\frac{d\varphi}{dz}\right)\right]-k_1^2\frac{d\varphi}{dz}\left(\frac{\partial v'}{\partial z}+u'\frac{d\varphi}{dz}\right), \quad \text{(IV.6)}$$

$$\frac{\partial^2 v'}{\partial t^2}=\frac{\partial}{\partial z}\left[k_1^2\left(\frac{\partial v'}{\partial z}+u'\frac{d\varphi}{dz}\right)\right]+k_2^2\frac{d\varphi}{dz}\left(\frac{\partial u'}{\partial z}-v'\frac{d\varphi}{dz}\right). \quad \text{(IV.7)}$$

Equations (IV.6) and (IV.7) are the simplest form of Maxwell's equations when dependence on x and y is neglected. If k_1, k_2 and φ are independent of z, (IV.6) and (IV.7) reduce to (III.21) and (III.22).

We now seek a simplification, by reduction of order, analogous to (III.24) and (III.25). To this end we consider the possibility that

$$\frac{\partial u'}{\partial z}-v'\frac{d\varphi}{dz}=\mp\frac{1}{k_2}\frac{\partial u'}{\partial t}, \quad \text{(IV.8)}$$

$$\frac{\partial v'}{\partial z}+u'\frac{d\varphi}{dz}=\mp\frac{1}{k_1}\frac{\partial v'}{\partial t} \quad \text{(IV.9)}$$

may be solutions of (IV.6) and (IV.7). If we substitute (IV.8) and (IV.9) into (IV.6) and (IV.7) we find

$$\frac{\partial^2 u'}{\partial t^2}=\frac{\partial^2 u'}{\partial t^2}\pm(k_1-k_2)\frac{d\varphi}{dz}\frac{\partial v'}{\partial t}\mp\frac{dk_2}{dz}\frac{\partial u'}{\partial t}, \quad \text{(IV.10)}$$

$$\frac{\partial^2 v'}{\partial t^2}=\frac{\partial^2 v'}{\partial t^2}\pm(k_1-k_2)\frac{d\varphi}{dz}\frac{\partial u'}{\partial t}\mp\frac{dk_1}{dz}\frac{\partial v'}{\partial t}. \quad \text{(IV.11)}$$

Equations (IV.10) and (IV.11) fail to be identities by terms whose coefficients have ratios, to the coefficients of the leading terms, of order of magnitude

$$\lambda\frac{dn}{dz} \quad \text{and} \quad [(n_1-n_2)/n]\lambda\frac{d\varphi}{dz}. \quad \text{(IV.12)}$$

The first of these has been shown, in Section III, to be of order of magnitude 10^{-7}. Hence, if the product of $(n_1-n_2)/n$ and $\lambda d\varphi/dz$ is negligible, (IV.6) and (IV.7) may be replaced by (IV.8) and (IV.9). In a material with a stress-optical coefficient of 100 Brewsters under a stress of 10^8 dynes/cm^2, the order of magnitude of $(n_1-n_2)/n$ is 10^{-3} while a rotation of the secondary principal axes in the amount 2π radians per millimeter makes $\lambda d\varphi/dz$ of order of magnitude 10^{-3}.

It is not impossible for $d\varphi/dz$ to become very large. An example is the case of combined tension and torsion of a cylindrical bar. For a wave-normal intersecting the axis of the bar, $d\varphi/dz$ approaches infinity as the tension approaches zero. However, at the same time $(n_1-n_2)/n$ approaches zero and the product of the two has a zero limit, in the neighborhood of the axis of the bar.

As a result of the considerations in this and the preceding section, we conclude that the propagation of a forward moving wave in a

photoelastic body under a three-dimensional state of stress is governed, to a close approximation, by the differential equations

$$\frac{\partial u'}{\partial z}-v'\frac{d\varphi}{dz}=-\frac{1}{k_2}\frac{\partial u'}{\partial t}, \tag{IV.13}$$

$$\frac{\partial v'}{\partial z}+u'\frac{d\varphi}{dz}=-\frac{1}{k_1}\frac{\partial v'}{\partial t}. \tag{IV.14}$$

V. PASSAGE TO NEUMANN'S EQUATIONS

Consider solutions of (IV.13) and (IV.14) of the form

$$u'=U\cos(\omega t-\alpha), \tag{V.1}$$

$$v'=V\cos(\omega t-\beta), \tag{V.2}$$

where the amplitudes U and V and the phase angles α and β are functions of z. Substituting these expressions for u' and v' in (IV.13) and (IV.14) and equating coefficients of like trigonometric terms of argument ωt, we find

$$U\frac{d\alpha}{dz}\cos\alpha+\frac{dU}{dz}\sin\alpha-V\frac{d\varphi}{dz}\sin\beta=\frac{\omega U}{k_2}\cos\alpha, \tag{V.3}$$

$$-U\frac{d\alpha}{dz}\sin\alpha+\frac{dU}{dz}\cos\alpha-V\frac{d\varphi}{dz}\cos\beta=-\frac{\omega U}{k_2}\sin\alpha, \tag{V.4}$$

$$V\frac{d\beta}{dz}\cos\beta+\frac{dV}{dz}\sin\beta+U\frac{d\varphi}{dz}\sin\alpha=\frac{\omega V}{k_1}\cos\beta, \tag{V.5}$$

$$-V\frac{d\beta}{dz}\sin\beta+\frac{dV}{dz}\cos\beta+U\frac{d\varphi}{dz}\cos\alpha=-\frac{\omega V}{k_1}\sin\beta. \tag{V.6}$$

Now, multiply (V.3) by $\sin\alpha$ and (V.4) by $\cos\alpha$ and add; multiply (V.5) by $\sin\beta$ and (V.6) by $\cos\beta$ and add; multiply (V.3) by $\cos\alpha$ and (V.4) by $\sin\alpha$ and subtract; multiply (V.5) by $\cos\beta$ and (V.6) by $\sin\beta$ and subtract. The following four equations result:

$$\frac{dU}{dz}=V\frac{d\varphi}{dz}\cos\delta, \tag{V.7}$$

$$\frac{dV}{dz}=-U\frac{d\varphi}{dz}\cos\delta, \tag{V.8}$$

$$U\frac{d\alpha}{dz}+V\frac{d\varphi}{dz}\sin\delta=\frac{\omega U}{k_2}, \tag{V.9}$$

$$V\frac{d\beta}{dz}+U\frac{d\varphi}{dz}\sin\delta=\frac{\omega V}{k_1}, \tag{V.10}$$

where δ is the relative phase retardation:

$$\delta=\alpha-\beta. \tag{V.11}$$

Subtracting (V.10) from (V.9) we have

$$\frac{d\delta}{dz}+\left(\frac{V}{U}-\frac{U}{V}\right)\frac{d\varphi}{dz}\sin\delta=-\omega\left(\frac{1}{k_1}-\frac{1}{k_2}\right)=\frac{-2\pi}{\lambda}(n_1-n_2). \tag{V.12}$$

Now, let the amplitude ratio be given by

$$V/U=\tan\gamma, \tag{V.13}$$

so that

$$(V/U)-(U/V)=-2\cot 2\gamma, \tag{V.14}$$

and

$$\frac{d\gamma}{dz}=\left(U\frac{dV}{dz}-V\frac{dU}{dz}\right)\Big/(U^2+V^2). \tag{V.15}$$

Substituting into (V.15) the values of dU/dz and dV/dz given in (V.7) and (V.8) we find

$$d\gamma/dz=-(d\varphi/dz)\cos\delta \tag{V.16}$$

and, substituting (V.14) into (V.12), we find

$$\frac{d\delta}{dz}+\frac{2\pi}{\lambda}(n_1-n_2)=2\frac{d\varphi}{dz}\cot 2\gamma\sin\delta. \tag{V.17}$$

Equations (V.16) and (V.17) are Neumann's equations,[5] derived in 1841 from purely kinematical considerations. Alternative derivations of Neumann's equations, again from kinematical considerations, are given by Coker and Filon[6]

[5] F. E. Neumann, Abh. d. Kön. Akad. d. Wissenschaften zu Berlin (1841) Part II, p. 1.
[6] See reference 3, p. 253.

and Kammerer.[7] These equations appear to have been used but little. Coker and Filon[6] employed them to estimate the effect on δ and γ of the quadratic variation of optical properties through the thickness of a plate under a state of plane stress. Menges[8] devised a graphical method of solving for (n_1-n_2) and φ from data obtained by measurement of δ and γ.

VI. REDUCTION TO A CANONICAL FORM

The contributions of the various optical properties of the medium are revealed by discovering the characteristic curves of (IV.13) and (IV.14) and then reducing the equations to a canonical form.

If we solve (IV.13) for v', differentiate and substitute in (IV.14), the result is a second order partial differential equation in one dependent variable, u', and two independent variables, z and t:

$$\frac{\partial^2 u'}{\partial z^2}+(k_2^{-1}+k_1^{-1})\frac{\partial^2 u'}{\partial z\partial t}+(k_2^{-1}k_1^{-1})\frac{\partial^2 u'}{\partial t^2}+\left(\frac{d^2\varphi/dz^2}{d\varphi/dz}\right)\frac{\partial u'}{\partial z}+\left(\frac{dk_2^{-1}}{dz}+k_2^{-1}\frac{d^2\varphi/dz^2}{d\varphi/dz}\right)\frac{\partial u'}{\partial t}+u'\left(\frac{d\varphi}{dz}\right)^2=0. \quad \text{(VI.1)}$$

The quadratic form associated with this equation is

$$Q(p)=ap^2+2bp+c=p^2+(k_2^{-1}+k_1^{-1})p+k_2^{-1}k_1^{-1}=(p+k_2^{-1})(p+k_1^{-1}). \quad \text{(VI.2)}$$

Since the discriminant, $ac-b^2$, of this quadratic form is always negative, Eq. (VI.1) is of the hyperbolic type. The differential equations which determine the characteristic curves are

$$\frac{\partial\sigma}{\partial z}=-k_1^{-1}\frac{\partial\sigma}{\partial t} \quad \text{and} \quad \frac{\partial\tau}{\partial z}=-k_2^{-1}\frac{\partial\tau}{\partial t}, \quad \text{(VI.3)}$$

whence the characteristic curves for Eq. (VI.1), and also for its counterpart obtained by eliminating u', are given by the constant values of

$$\sigma=\omega\left(t-\int k_1^{-1}dz\right), \quad \text{(VI.4)}$$

$$\tau=\omega\left(t-\int k_2^{-1}dz\right). \quad \text{(VI.5)}$$

These reveal the manner in which the wave solutions of (IV.13) and (IV.14) are propagated along the positive z-direction.

Transforming from variables z and t to σ and τ and defining

$$R=\frac{d\varphi/dz}{\omega[(1/k_2)-(1/k_1)]}=\frac{d\varphi/dz}{(2\pi/\lambda)(n_2-n_1)}, \quad \text{(VI.6)}$$

Eqs. (IV.13) and (IV.14) become, finally,

$$Ru'=\partial v'/\partial\tau, \quad \text{(VI.7)}$$

$$Rv'=\partial u'/\partial\sigma. \quad \text{(VI.8)}$$

The quantity R is thus shown to be the essential parameter controlling optical wave propagation in three-dimensional photoelasticity. R is a ratio whose numerator is the rate of rotation of the secondary principal axes and whose denominator is the rate of relative phase retardation that would obtain in the absence of rotation. This ratio is a generalization of a similar quantity discovered in the solution of Maxwell's equations for the case in which both $d\varphi/dz$ and n_2-n_1 are constants.[2] In that solution it was found that the influence, on observed relative phase retardation, of the uniform rotation of the secondary principal axes could be expressed entirely in terms of R. It now appears that this must be true regardless of the manner in which both $d\varphi/dz$ and n_2-n_1 vary. The specific functional dependence of relative phase retardation on R will, of course, depend on the form of R itself. It would be useful to discover the influence of R in a number of special cases, a process which now appears feasible upon noticing the relative simplicity of Eqs. (VI.7) and (VI.8) in comparison with Maxwell's or Neumann's equations. In particular it is to be noted that, if u' and v' are separated between (VI.7) and (VI.8), the equation governing each of them is a special form of the general one-dimensional wave equation.

[7] A. Kammerer, *Recherches sur la photoélasticimétrie* (Hermann et Cie., Paris, 1944), p. 37.

[8] H. J. Menges, Zeits. f. Angewandte Math. u. Mech. 20, 210 (1940).

A Method for Measuring the Total Power of Small-Angle X-Ray Scattering*

B. E. WARREN
Eastman Laboratory of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received July 27, 1948)

Measurements of the absorption coefficients are made with a double crystal spectrometer in the parallel position, placing the sample between the two crystals, and between the second crystal and the chamber. In the first position the beam is cut down by the ordinary absorption and also by the power going into small-angle scattering; in the second position it is cut down only by the ordinary absorption. The difference in the two measured absorption coefficients gives directly the total power in small-angle scattering. From the latter, an average spherical particle size can be computed.

MATERIALS in which there is an appreciable variation in the density of scattering power over distances of the order of 20–1000A give rise to a small-angle x-ray scattering, which is readily observed. The techniques in common use are designed to measure the intensity of small-angle scattering in arbitrary units as a function of the angle of scattering. The method described in this paper gives the total power of small-angle scattering in absolute units, that is, the ratio of the total power per gram of scattering substance to the intensity of the primary beam.

METHOD FOR MEASURING THE TOTAL POWER

A double crystal spectrometer[1] equipped with good calcite crystals is set for maximum intensity in the parallel position. The chamber (or Geiger counter) slit is made rather broad. The small-angle scattering substance is used in the form of a thin slab of optimum thickness. It is placed in either position A between the two crystals, or position B between the second crystal and the chamber slit. With the sample in position B, the small-angle scattering passes through the broad chamber slit, and the intensity of the transmitted beam, as read by the Geiger counter, is cut down by an amount corresponding to the ordinary absorption coefficient of the material. With the sample in position A, the major part of the small-angle scattering is sufficiently deviated so as not to be reflected by the second crystal, and the transmitted beam, as recorded by the Geiger counter, is cut down by both the ordinary absorption effect and the loss of intensity by small-angle scattering. Let m be the mass per cm² of scattering substance, μ_m the ordinary mass absorption coefficient of the substance, and μ_{SA} the additional mass absorption coefficient resulting from small-angle scattering.

$$\begin{aligned} I_A &= I_0 \exp(-(\mu_m+\mu_{SA})m), \\ I_B &= I_0 \exp(-\mu_m m), \\ \mu_{SA} &= 1/m \ln(I_B/I_A) \\ &= \mu_m((\ln I_B/I_A)/(\ln I_0/I_B)). \end{aligned} \tag{1}$$

Let P_{SA} be the total power in small-angle scattering per unit mass of scatterer assumed to be irradiated by the constant intensity I. The decrease in intensity of the transmitted beam in a thin slab of mass per unit area dm caused by small-angle scattering only is given by

$$(dI)_{SA} = -I\mu_{SA}dm = -P_{SA}dm, \tag{2}$$

$$\begin{aligned} P_{SA}/I = \mu_{SA} &= 1/m \ln(I_B/I_A) \\ &= \mu_m((\ln I_B/I_A)/(\ln I_0/I_B)). \end{aligned} \tag{3}$$

From the relative intensities I_0, I_A, and I_B, and either the mass per unit area of the sample or the ordinary mass absorption coefficient of the sample, we get directly the ratio of the total power in small-angle scattering per unit mass of sample to the intensity of the primary beam.

APPLICATION TO A SUBSTANCE COMPOSED OF SPHERICAL PARTICLES

Assume that the sample is composed of particles that are spherical and all of the same size, with a sufficiently random center to center distance to allow adding the intensities of small-

* Presented at the Summer Meeting of the A.S.X.R.E.D., St. Marguerite Station, Quebec, June 25, 1947.

[1] A method for using the double crystal spectrometer to measure the intensity, rather than the total power of small-angle scattering, has been described by I. Fankuchen and M. H. Jellinek, Phys. Rev. 67, 201 (1945).

angle scattering per particle. In general, with solid samples this assumption of incoherency in the scattering per particle is probably not well fulfilled; it is worse the smaller the angle of scattering. The intensity of small-angle scattering per gram of sample is given by[2]

$$I_{SA} = I_e n N^2 [(3)/((kr)^2)]\{(\sin kr)/(kr) - \cos kr\}]^2$$
$$= I_e n N^2 (9\pi)/(2(kr)^3) J_{3/2}^2(kr), \quad (4)$$

where r = the particle radius, N = the number of electrons per particle, n = the number of particles per gram, $k = (4\pi \sin\theta)/(\lambda)$, and $I_e = I(e^4)/(m^2c^4R^2)$.

The total power of small-angle scattering is given by integrating I_{SA} over the area where $dA = 2\pi R^2 \sin 2\theta d(2\theta)$,

$$P_{SA} = \int I_{SA} dA = I_e R^2 n (N^2 9\lambda^2)/(4r^2) \times \int_0^\infty (1/(kr)^2) J_{3/2}^2(kr) d(kr), \quad (5)$$

$$P_{SA} = I(e^4 n N^2 9\lambda^2)/(m^2c^4 8\pi r^2),$$

$$P_{SA}/I = \mu_{SA} = (e^4 n N^2 9\lambda^2)/(m^2c^4 8\pi r^2).$$

Let ρ be the density within a particle, L the Avogadro number, Z the atomic number, A the atomic weight, and $\sum$ represent summation over some unit of composition. The number of electrons per particle, and the number of particles per gram, can be expressed as follows:

$$N = (4/3)\pi r^3 \rho L(\textstyle\sum Z)/(\sum A),$$
$$n = (1)/[(4/3)\pi r^3 \rho].$$

Substituting in (5), we obtain

$$\mu_{SA} = \tfrac{3}{2}(e^4)/(m^2c^4) L^2 \lambda^2 ((\textstyle\sum Z)/(\sum A))^2 \rho r. \quad (6)$$

For low atomic numbers we can set $(\sum Z)/(\sum A) = \frac{1}{2}$. Expressing λ and r in Angstrom units, we obtain, finally,

$$\mu_{SA} = 0.0108 \lambda_A^2 \rho r_A. \quad (7)$$

According to Eq. (7), the additional absorption resulting from small-angle scattering increases proportional to the size of the spherical particles. The equation breaks down, however, for particles so large that the small-angle scattering is confined to such small angles that a major part of it is passed by the second crystal.

The following data, obtained from a sample of carbon black, illustrate the application of Eq. (7). The sample was Godfrey L. Cabot Grade 6 Channel Black, packed loosely in a sample container 1.6 mm thick. The radiation was Cu $K\alpha\lambda = 1.54$A; for this wave-length the mass absorption of carbon is $\mu_m = 4.52$ cm^2/g. The three intensity readings in arbitrary units were $I_0 = 42.0$, $I_A = 12.5$, $I_B = 29.5$.

$$\mu_{SA} = 4.52 \ln(29.5)/(12.5)/\ln(42.0)/(29.5)$$
$$= 11.1 \text{ cm}^2/\text{g}. \quad (8)$$

Notice that the additional absorption caused by small-angle scattering is for this material more than twice the ordinary absorption. Assuming the density within a carbon black particle to be about $\rho = 2.0$, the particle radius can be computed from Eq. (7):

$$r_A = (11.1)/(0.0108 \times (1.54)^2 \times 2.0) = 220\text{A}.$$

Previous measurements[2] of the intensity of small-angle scattering from the same material gave a size distribution in which the radius ran from 95–275A.

The method outlined in this paper is simple and quick to apply since it involves making only three relative intensity measurements. It might conceivably have special use in putting an intensity curve obtained in the usual way upon an absolute basis. Since this method delivers only one average quantity, it cannot compare with the usual small-angle scattering methods for studying the shape of particles or a particle size distribution.

It is a pleasure to acknowledge the financial assistance by Godfrey L. Cabot, Inc., which has made possible a program of small-angle scattering studies, and to express my thanks to Dr. W. R. Smith, Chief Research Chemist of Godfrey L. Cabot, Inc., for his generous help and advice.

[2] J. Biscoe and B. E. Warren, J. App. Phys. **13**, 364 (1942).

The Electron-Micrographic Structure of Shadow-Cast Films and Surfaces

ROBLEY C. WILLIAMS AND ROBERT C. BACKUS
Randall Laboratory of Physics, University of Michigan, Ann Arbor, Michigan
(Received June 28, 1948)

The lower limit of size of biological objects which can be photographed with the electron microscope, by the aid of shadow-casting, is shown to depend upon the smoothness of substrate upon which they can be mounted, and upon the continuity of structure of the thin films with which they are shadowed. Numerous attempts to improve the existing deficiencies are reported, both with respect to producing smoother substrate films, and to producing films for shadow-casting of high efficiency and continuity of structure.

No success has been encountered in producing a usable substrate film perceptibly smoother than the collodion and Formvar films commonly used. It is found that the best shadow-casting material for this type of film is uranium or uranium oxide.

Verification has been obtained of the severe granulation of gold films previously used in the pre-shadowed replica process, when subjected to the electron current of a biased-beam electron gun. Attempts to reduce the granulation to a satisfactory level have failed. Various methods of preparing pre-shadowed replicas are reported. The factors affecting the tenacity of evaporated films to glass surfaces are discussed. It is found that elements which oxidize readily are relatively adherent to glass, while the elements gold, palladium, platinum, and rhodium are not. Uranium sulfide can be used as a pre-shadowed replica material, but only with some uncertainty, owing to its chemical instability. It has been found that a palladium-platinum mixture is the most satisfactory material for use in the pre-shadowed replica technique, and that films of this mixture in a thickness of about 6A produce adequate shadows in which there is no sign of granulation. The surface of clean glass is again found to have the smoothest structure of any material known, with practically no sharp discontinuities in elevation as great as 10A.

An appendix is given, in which technical details of shadow-casting and replica production are described.

INTRODUCTION

IN 1944 Williams and Wyckoff[1] demonstrated that the electron micrography of minute biological objects could be greatly improved by the application of a technique which has come to be called "shadow-casting" or "shadowing." Very small particles of substances of organic nature, such as are found in biological material, ordinarily do not have sufficient electron scattering power to make their detection possible when they are mounted on a collodion film substrate in the usual manner of specimen preparation. Consequently, prior to the advent of the shadowing technique, the electron microscope could not be utilized to the full extent of its inherent resolving power in the examination of small, particulate biological material, and, in fact, it was generally accepted that a particle diameter of about 200A was the lower limit of detectability. This is in marked contrast to the minimum discernible diameter of particles of a substance of high atomic number, such as a gold sol, where a size on the order of 10A has been reported.[2]

In the shadow-casting process the specimen is coated, in vacuum, by the oblique deposition of a layer of a material relatively opaque to electrons, such as chromium. As a result any particle which is elevated above the general layer of the substrate film prevents the formation of the chromium layer in the immediately adjacent region opposite to the direction of the source of chromium, creating an area which might be termed a metallic "shadow." When the shadowed specimen is photographed in the electron microscope, the areas relatively devoid of the metallic film are more transparent to the electrons than the adjacent areas of the substrate, and hence appear dark on the photographic negative, while the object casting the shadow appears relatively white. The general effect is the same as that obtained in a positive photograph of a region reflecting light under oblique illumination.

Williams and Wyckoff were able to photograph with considerable clarity objects of organic nature having diameters of approximately 100A, by the use of gold as the shadowing agent, and collodion as a substrate film. They found, however, that this technique was not applicable in its simple form to the detection of objects

[1] R. C. Williams and R. W. G. Wyckoff, J. App. Phys. **15**, 712 (1944).
[2] James Hillier, J. App. Phys. **17**, 307 (1946).

smaller than 100A, owing to the lack of smoothness of the surface of the collodion film, which has a "pebbly" structure of the order of 50A in size. They developed a technique which they called a "pre-shadowed replica" method,[3] in which the shadowing is first done with the specimen material mounted on a glass surface. Subsequently the specimen and shadowing layer are stripped from the glass by the use of a collodion film, and effectively are transferred to the surface of that film. This procedure was very satisfactory at the time, since the surface of polished glass was found to be extremely smooth, and the inherent surface structure of the collodion film was not exhibited because it was not shadowed.

Fairly recently there has been developed for use in the electron microscope a biased-beam electron gun[4] which has a desirable characteristic of producing an intense electron beam in the plane of the specimen. However, the intense electron beam from this type of gun has been reported[5] to produce an agglomeration of the gold film used in shadowing, with the result that the film appears broken into minute globules; an effect commonly called "granulation." The advantages of the biased beam are too great to consider abandoning it in order to preserve the usefulness of gold films, and hence the problem of suitable shadowing agents and specimen substrates has had to be examined anew. In particular, our problem has been to develop methods whereby the presence and approximate shapes of biological particles of molecular weights around 150,000–200,000 (one dimension less than 40A) can be ascertained.

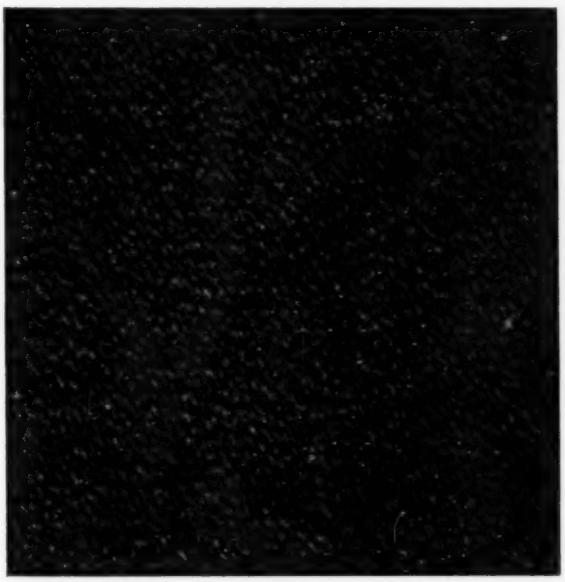


FIG. 1. A water-spread surface of collodian, shadowed with a film of uranium about 6A thick at a grazing angle of 11° (a "5 to 1" angle). Scale mark represents 0.1 micron.

[3] R. C. Williams and R. W. G. Wyckoff, Science **101**, 594 (1945).
[4] James Hillier and R. F. Baker, J. App. Phys. **16**, 469 (1945).
[5] R. J. Mandle, Proc. Soc. Exp. Biol. Med. **64**, 362 (1947).

If only a sufficiently smooth substrate film existed, the problem of photographing particles of diameters less than 50A would be solved very simply. Materials exist for shadowing which do not granulate (as will be discussed later), and they are of such high atomic number and density that films only 4–6A thick, as measured normally to the substrate, will cast shadows of sufficient contrast to be usable. On the other hand, the pre-shadowed replica method alone would be a good solution, since the surface of clean glass is adequately smooth, if only there could be found substances for shadowing which did not granulate and could be readily stripped from the glass to form the necessary replicas.

Our attempts to solve the problem have taken two obvious paths: (1) the search for a smooth substrate film to be used in direct shadowing and (2) the search for a proper shadowing material to be used in making pre-shadowed replicas.

II. ATTEMPTS TO IMPROVE THE SURFACE SMOOTHNESS OF SUBSTRATE FILMS TO BE USED FOR DIRECT SHADOW-CASTING

A. Shadow-Casting Substances Which Do Not Granulate

Collodion or Formvar (polyvinal formal) films are commonly made for use in electron microscopy by first forming the film on water, and transferring this to the fine-mesh microscope screen. As has been indicated earlier, such films are entirely satisfactory for biological objects of particulate nature, if they have diameters greater than about 100A, and for relatively large bio-

logical objects whose surface detail is the main point of interest.

If fine detail in the specimen is to be observed, however, three conditions are imposed upon the shadowing agent. It should have a high scattering power for electrons, so as to be usable in a very thin layer; it should be readily evaporated in the vacuum chamber used for shadow-casting; it should not granulate under the impact of the electron beam. As several investigators have shown, the elements of highest melting point are least likely to granulate, and the obvious material to use to satisfy the first and last conditions mentioned above is uranium.[6] Its only objectionable characteristic is the relative difficulty with which it is evaporated, since it alloys readily with the tungsten filament.

We have reconsidered the use of uranium, and find, in agreement with others, that a thin uranium film is completely oxidized upon any exposure to air, and certainly by the time it is examined in the electron microscope. It seems entirely reasonable, then, to suppose that uranium oxide (U_3O_8) is a satisfactory material to evaporate in the first place, and we have found it to be so, particularly inasmuch as it does not perceptibly attack the tungsten filament. When deposited in a thickness calculated to be only 6A, it produces faint, but discernible, shadows. Tungstic oxide can also be readily evaporated, and makes a somewhat less satisfactory shadowing agent than uranium oxide, owing to its lower average atomic number. A disadvantage of the use of uranium oxide is its very high vaporizing point, resulting in the possibility of damage to the specimen if it is placed too near the filament during shadow-casting.

B. Attempts to Reduce the Irregularities of the Substrate Film

Both collodion and Formvar are commonly used as substrate films, but inasmuch as we have never detected any systematic difference in their surface textures, we shall refer mostly to collodion with the understanding that the same conclusions apply to Formvar. The structure of a collodion film (U.S.P., diluted about 10-fold with amyl acetate) spread on water, is shown in

[6] R. C. Williams and R. W. G. Wyckoff, J. App. Phys. 15, 712 (1944).

FIG. 2. The characteristically rough surface of a thick film deposited in vacuum. The example shown in this micrograph is that of a collodian replica of a film of aluminum 1000A thick. Replica shadowed with uranium at a 5 to 1 angle. Scale mark represents 1 micron.

Fig. 1. The surface was shadowed with about 6A of uranium, at a 5 to 1 angle. The considerable irregularity of its surface is evident, and it is clear that particulate objects around 50A in diameter that might exist in a preparation would not be detectable when masked by the structure of the collodion background. We have compared the air-exposed surface of spread films of collodion with the water-exposed surface, and find no difference in structure.

a. We have prepared collodion films by allowing them to dry from the solvent while in contact with a smooth solid surface, such as glass. The film can then be stripped off and the side which has dried in contact with the glass subsequently shadowed. This method has been tried with both collodion and Formvar, and with amyl acetate, dioxane, and ethylene dichloride as solvents, but the surface structure always has been found to have irregularities of the same magnitude as shown in Fig. 1. Evidently the highly polymerized units in both types of film retain their identity, form, and interwoven structure following the evaporation of the solvent.

b. Another possible method of reducing the surface structure of the supporting film is to coat

the collodion film with another film whose inherent surface structure might be expected to be finer, with the hope that the collodion structure would be filled in or bridged over. A promising material for this purpose is a film of silicon monoxide deposited in vacuum. It is known that films of SiO are relatively structureless, as observed by electron diffraction, and that they are sufficiently strong to serve as substrate films when only 20A thick.[7] We have found however, that when films of SiO are deposited upon collodion, in calculated thicknesses varying from 20 to 100A, the structure of the collodion surface is imparted to the SiO layer, and no smoothing of the former is effected.

c. There is no inherent reason why the substrate film must be formed of collodion or Formvar, and some work has been done with films of other materials. Ellis[8] has reported on the surface structure of a low polymer of polyethylene, shadowed with chromium, and has found rather gross irregularities. We have tried polyethylene dissolved in warm xylene and cast as a film on hot water, and also flowed as a film upon hot, smooth glass. In both cases the surface irregularities, as photographed subsequent to shadowing with uranium, have been found to be larger than those of collodion.

d. We have experimented with the use of a film of SiO deposited directly upon glass, in order to ascertain whether or not the glass-deposited surface is sufficiently smooth to be useful. However, a film of SiO cannot readily be stripped from clean glass, and we have tried pre-coating the glass with very thin films of soluble materials such as sodium chloride, boron, and copper to make possible the removal of the SiO film. In every case we have found that, if the sub-film is thick enough to allow the SiO to be removed, it is then thick enough to introduce into the SiO surface the characteristically rough structure of a thick film deposited in vacuum (Fig. 2).

III. IMPROVEMENT OF PRE-SHADOWED REPLICAS

A. The Granulation of Gold

We have performed some experiments with gold-shadowed replicas, to assure ourselves of the conditions under which the granulation of the gold occurs. We find that these conditions are: too great a thickness of the gold film, deposition of the film in a poor vacuum, and, most importantly, an electron beam current of too high intensity. It is our experience, in accord with that of others, that *if an electron microscope equipped with a biased-beam gun is to be used for the examination of minute structures, the use of gold as a shadowing agent is inadvisable.*

B. Experiments on the Possible Reduction of Gold Granulation

Gold is a very convenient shadowing material, easy to evaporate, of high atomic number, and readily stripped from the glass. We have felt that its use should not be abandoned without investigating the possibility that its granulation could be decreased below a troublesome value. Two possible approaches are evident: to attempt to stabilize or fix the gold film before it has been stripped from the glass by the collodion film, and to attempt to stabilize it subsequent to stripping but prior to insertion in the microscope.

Whether or not these procedures appear naïve hinges upon one's conception of what occurs when an originally continuous metallic film aggregates into droplets at temperatures several hundred degrees below its melting point. The phenomenon is apparently one of atomic diffusion, but there is complete ignorance as to what might serve as centers upon which the small aggregates form. An analogous phenomenon is the "sintering" or aggregation of a gold black at a temperature some 800°C below the melting point of bulk gold. It has occurred to us that a substance which would interfere with the free diffusion of the gold atoms, or would inhibit the beginnings of the formation of aggregates, might prevent the customary granulation.

a. In an attempt to stabilize the gold film before stripping, two materials of high melting point, SiO and uranium, were deposited along with the gold. Films of these were applied before, after, and during the deposition of the gold. When they were deposited *before*, and if made thin enough to allow the gold to be stripped off, they did not prevent granulation. *Mixed* with the gold film they appeared to have no effect, at least until made such a high percentage of the

[7] C. E. Hall, J. App. Phys. **19**, 198 (1948).
[8] S. G. Ellis, J. App. Phys. **18**, 846 (1947).

TABLE I.

Element deposited	Ease of removal
Beryllium	−
Aluminum	−
Chromium	−
Manganese	−
Nickel	±
Copper	±
Rhodium	+
Palladium	+
Silver	±
Indium	±
Tungsten	−
Platinum	+
Gold	+
Lead	−
Uranium	−

+, Readily; ±, occasionally; −, non-removable.

total film that it could not be stripped. When deposited *after* the gold, they had no effect upon granulation.

An attempt was made to avoid the use of collodion or Formvar in the stripping process, and to use, instead, a thick film (100–200A) of SiO for the supporting membrane. This type of replica was prepared, but no improvement in gold granulation was found.

b. Attempts to stabilize the gold film subsequent to stripping were based upon the hypothesis that the inevitable distortion of the collodion upon electron impact might produce an incipient breaking of the gold film which would lead to agglomeration. Films of SiO were deposited upon either one or both sides of the collodion film, after the gold film replica had been stripped, and although it was found that the *collodion* film was stabilized thereby, the gold aggregated into droplets as before.

In conclusion to this set of trials, we can say that the only feasible means of reducing gold granulation appears to be the well-known procedure of decreasing the intensity of the electron beam.

C. Experiments with Pre-Shadowing Agents Other Than Gold

a. Factors Affecting the Adherence of Films to Glass Substrates

A material satisfactory for use in forming pre-shadowed replicas from a glass surface must be of high electron scattering-power, of non-granulating character, and should be readily removable from the glass by the ordinary collodion stripping technique. In addition, it is convenient if the material is readily evaporated from a tungsten filament in an ordinary shadow-casting equipment. The first requirement restricts us to the use of elements, or compounds, of average atomic number around 60 or greater, while the second requirement apparently restricts us to the use of materials of high melting point, above 1500°C. The nature of the restrictions consequent upon the third requirement (removability of the film) is not very well understood, and most of our efforts have been spent in elucidating this point further.

The relative adherence of evaporated films of metal to glass substrates has been for many years a problem of great practical importance and of extreme empiricism. One of us[9] had found in 1934 that a film of a relatively non-adherent metal, such as gold, could be rendered strongly adherent to glass by the pre-application of a thin film of chromium or beryllium, but the cause of this effect is still obscure. We have made some systematic observations of the tenacity of metal films to glass, and have the following general conclusions to offer.

Gold will normally adhere very weakly to glass, and a gentle pull applied to a film of collodion, dried in contact upon it, will strip it from the glass. The character of the glass surface prior to the deposition of the gold is important, however, and we have found that the surface can be prepared in ways such as to cause the gold to adhere quite firmly. One way is to clean the glass strongly with a chromic-acid cleaning solution; this probably etches the glass surface and creates crevices in which the gold is mechanically anchored. Another way is to cause the surface of the glass to be strongly hydrophobic at the time the gold is deposited. This can result from the presence of a thin film of oil on the surface (applied by rubbing, or even by using an inadequately trapped diffusion pump operating with Silicone oil), or by ionic bombardment within the vacuum chamber. Whatever the means of destroying the natural hydrophilic surface character of clean glass, we have made

[9] R. C. Williams, Phys. Rev. 46, 146 (1934).

the general observations that gold (and other ordinarily non-adhering elements) will adhere to glass if its surface is not readily wet with water.

In speculating upon the causes of this effect, it should be remembered that no metal film can be stripped from glass with the aid of collodion unless the combination has been thoroughly moistened, usually with condensed breath moisture. Possibly it is necessary that water molecules penetrate the collodion, and enter as a film between the metal and the glass in order that the former can be stripped, and if the glass surface is hydrophobic the entry of water will not occur. This hypothesis is supported by the observation that a gold film, coated with film of SiO (which is relatively impervious to water in a thickness of 100A), can be stripped only with difficulty.

b.

We have experimented with several materials other than gold in an attempt to discover what characteristics of the material determine its relative adherence to glass, when deposited as a film about 20A thick. Most of our work has been with elements, rather than compounds, and we will report results with these first.

Thin films of approximately 20A thickness were deposited upon clean* microscope-slide glass, and their relative adherence tested by their resistance to removal by means of stripping with a film of collodion. The results following in Table I were obtained.

These data reveal only one reasonably consistent correlation between the character of the element and the relative tenacity; those elements which readily oxidize on exposure to air adhere strongly, while those which do not oxidize are readily removed from the glass. There appears to be no correlation with atomic number, hardness, density, melting point, or crystal structure of the elements. If the generalization is correct, it leads one to anticipate that films of vacuously deposited oxides would be strongly adherent to the glass. We have tested this prediction with the oxides of aluminum, silicon, tungsten, lead and uranium, and have found all of them tightly adherent.

* The glass microscope slides were cleaned by first scrubbing them vigorously with Dreft, followed by thorough rinsing with distilled water. The wet, rinsed slides were then held in steam issuing from a water still. This heated them to the point that they dried almost instantly upon removal from the steam.

TABLE II.

Element	Melting point	Atomic number	Specific gravity	Evaporation characteristic	Granulation
Pd	1555	46	12.0	Simple	Very slight
Pt	1755	78	21.4	Difficult	None apparent
Rh	1955	45	12.5	Very difficult	None apparent

The reasons for this selective behavior can be only surmised at this time. We believe that the effect of water or water vapor is an important and ever-present one, since collodion and Formvar films alone cannot readily be stripped from clean glass without having been moistened with condensed water vapor. It is conceivable that a monofilm of water is absorbed between the metal or metal-oxide film and the glass (mostly SiO_2), and that in this case hydration occurs with the formation of a hydrogen bond between the oxide layers.

c. Results with Removable Films

Our work indicates three conceivable solutions of the problem of producing films for use in pre-shadowed replicas which will meet the requirements listed under "*a*" of this section. The first is to find a smooth substrate from which even oxide films can be removed. (This would allow one to use uranium, for example, as a shadowing agent.) The second is to use a removable combination of oxidizing and non-oxidizing films. The third is to prepare only films which do not oxidize for use as the shadowing agents. We have experimented with all three solutions.

1. There appear to be very few natural or artificial surfaces which are satisfactory for use as pre-shadowing substrates, and are also sufficiently smooth for our purposes. Natural crystals, when cleaved, show relatively gross imperfections of structure, and when polished exhibit the polish marks of the abrasive. Materials which are even slightly soluble in water are generally unsuited for use with the aqueous solvents employed in biological investigations. Polished metals, without exception, exhibit polish marks, and also show their polycrystalline nature

when the debris formed by polishing is removed. Mica might appear to be a possibility, but we have found that its behavior with respect to deposited oxide films is like that of glass. In summary, we have found no surface, except glass, that is sufficiently smooth and insoluble to be used as a substate for forming pre-shadowed replicas.

2. It is conceivable that one can effect a combination of a film which is readily removable with one which is not, and utilize the removable feature of the former with the non-granulating feature of the latter. We have tried this possibility, using gold as the removable component, and the oxides of several elements as the non-granulating components. Without exception it is found that when the gold is present in sufficient quantity to allow the film to be removed the resulting combination film exhibits granulation.

3. The third possible solution has at least three variations, all of which we have tried. The first variation is to deposit a film of an oxidizable material, such as uranium, and to prevent its subsequent oxidation by protecting it, or by combining it with an element other than oxygen.

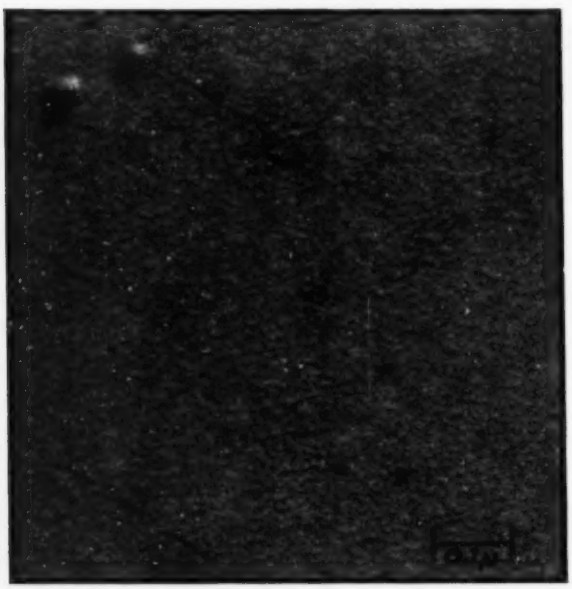


FIG. 3. A pre-shadowed replica of a glass microscope slide upon which palladium-platinum was deposited in a measured thickness of 6A at a shadowing angle of 5 to 1. Area shown in micrograph has more fine scratches than normally encountered. The depth of the scratch at the bottom of the picture is calculated to be about 20A. Magnification the same as in Fig. 1.

Protection was attempted in our work by admitting dry hydrogen to the vacuum chamber, subsequent to the uranium shadow-casting, and also by coating the uranium film, while still in vacuum, with a film of SiO. In both cases the uranium film oxidized, as demonstrated by its tenacity to glass, and by its electron diffraction pattern. The method of substitution for the prevention of the oxide film proved to be more promising. In this case the uranium film was either exposed to H_2S while still in the vacuum chamber, or was quickly dipped into a solution of H_2S in Formvar upon removal from the vacuum chamber, in an attempt to form a uranium sulfide film. The uranium film, so treated, proved to be frequently removable, although the results were not consistent.

The second variation is to deposit films of compounds which are not oxides, but which are stable upon vacuum evaporation, and do not subsequently change to oxides. Inasmuch as we are interested only in the high melting-point compounds of substances of high electron scattering-power, we have tried only the use of uranium sulfide. It evaporates with some difficulty, but it does not alloy with the tungsten filament. However, we have found its use somewhat uncertain, in that it can only occasionally be stripped from the glass substrate by the collodion replica technique.

The third variation is the obvious one using only elements which do not oxidize, and which are also satisfactory in other respects. We have made films of palladium, platinum and rhodium whose pertinent characteristics are indicated in Table II.

Platinum appears to be the most likely element, in respect to its high atomic number and density, and its freedom from granulation. The only limitations to its use are the very high temperature at which it evaporates in a vacuum, and the vigorous alloying action of molten platinum upon tungsten. We have attempted to reduce these difficulties by the use of a small admixture of palladium (about 20 percent by weight) with the platinum. The palladium melts relatively easily and quickly, and helps to melt the platinum by creating good thermal contact between filament and platinum. The net result is that the Pd-Pt alloy evaporates without the

usual spattering associated with the evaporation of pure platinum. It appears that a film of Pd-Pt is generally satisfactory for use with pre-shadowed replicas, and it is of interest to investigate the proper thickness of film to use. Shadowing should be done with a film of thickness sufficient only to establish adequate contrast in the shadows, since a thicker film produces a non-informative black-or-white micrograph and also results in distortion of the shapes of the shadowed particles. Naturally, what constitutes a shadow of minimum contrast is a matter of taste, and a function of the degree to which one is willing to enhance contrast by photographic means. It should be pointed out, however, that the production of vivid shadows by the use of a thick metal film is not consistent with undistorted micrography of particles of a size approaching the ultimate size resolvable by the electron microscope.

Rough calculation shows that a film of platinum should be as effective an electron scattering agent, per unit thickness, as is a film of uranium, since the latter always grows to the oxide with a consequent reduction of average atomic number and density. The effect of the palladium is to reduce the electron scattering power per unit thickness of the platinum, but in the relative amount used by us the effect of the dilution factor is less than 10 percent. We have measured the thickness of a Pd-Pt film just adequate for shadowing. Two microscope slides were placed in the vacuum chamber, one of which was located 15 cm from the filament, and whose surface made a $\tan^{-1}\frac{1}{5}$ with the surface and filament (commonly called a "5 to 1 angle"). The other slide was placed 10 cm from the filament, with a normal to its surface intersecting the filament. Thus the latter surface received about 10 times as thick a film as the former. Replicas were made of the inclined surface, and a micrograph of one is shown in Fig. 3. A collodion stripping was also made of the closer surface, covered by the thicker film, and the slide reinserted in the vacuum chamber. It was then shadowed at a 5 to 1 angle, and a collodion replica obtained. It was found that where there was an occasional patch of the thick Pd-Pt film which resisted the first stripping, it was coated upon re-shadowing the glass, and was then removed by the second stripping. A micrograph of such a patch, shows it to be shadowed, and by measuring the length of the shadow, one obtains the thickness of the patch. This thickness, divided by 10, gives the thickness of the film responsible for the shadows in Fig. 3. It is found that the thickness, so determined, is about 75 percent of that calculated from the geometry of the shadowing arrangement and the known quantity of Pd-Pt used. We conclude, that a calculated film thickness of 5A of Pd-Pt will yield perceptible shadows of small objects.

It is worth while to call attention again, as demonstrated by Williams and Wyckoff,[10] to the extreme smoothness of a glass surface. Our trials indicate that there are no perceptible differences in fine-scale smoothness between glass surfaces which are commercially rouge-polished or fire-polished, as long as the surfaces are not old enough to be atmospherically etched. The rouge-polished surfaces will show infrequent gross scratches, of course, of the order of 500A in depth. It appears that there are rarely sharp discontinuities in elevation as large as 5–10A on the surface of any freshly polished glass.

ACKNOWLEDGMENT

This work has been done as a preparatory phase of an electron micrographic study of the physical properties of large biological molecules. It was supported in large part by a grant from the American Cancer Society upon recommendation of the Committee on Growth of the National Research Council.

APPENDIX

1. The Shadow-Casting Process

Some descriptions of the shadow-casting process have been published[11] but no comprehensive and detailed discussion of the technique seems to exist.

a. The proper *substances* to use for shadowing have been discussed in this paper and elsewhere,[12] but in summary this can be said: For general purposes, where detail of structure of roughly 200A or larger is of interest, chromium or manganese are the most convenient and reliable elements to use. For the elucidation of contours of relatively large objects, such as bacteria and fibers, the shadowing films

[10] R. C. Williams and R. W. G. Wyckoff, Science **101**, 594 (1945).
[11] R. C. Williams and R. W. G. Wyckoff, J. App. Phys. **15**, 712 (1944), J. App. Phys. **17**, 23 (1946).
[12] W. T. Dempster and R. C. Williams, Anatomical Record **96**, 27 (1946).

should be applied at an angle of about 4 to 1, and in a computed thickness of about 60A. If cast upon a collodion substrate at angles more oblique than about 6 to 1, they will build up into a distinctly rough-appearing layer.

For fine detail of structures which are mounted on film substrates, or for the visual rendition of very small particles similarly mounted, the best materials appear to be uranium or uranium oxide. They will cast perceptible shadows in a computed thickness of about 4A, and intense shadows in a thickness of 8A. The most generally useful shadow-casting angle here is 5 to 1.

For pre-shadowed replicas, palladium-platinum or platinum alone appear to be most useful, in thicknesses about the same as for uranium. Gold is very convenient, and is usually satisfactory if not made thicker than about 8A and used in an instrument with a non-biased-beam electron gun. Granulation is made evident by the appearance on the micrographs of small globules which cast no shadows.

b. The calculated *thickness* of films is reliable, if an ordinary conical filament is used for evaporation and if an efficiency factor of about $\frac{3}{4}$ is allowed. The thickness is calculated to be that which exists on a plane surface making an angle α with the direction line to the filament. The formula for calculation is:

$$t=\frac{3}{4}\frac{M\cdot\tan\alpha\cdot 10^8}{4\pi r^2\cdot d},$$

where: t = thickness in Angstrom units, M = mass of material in grams, α = angle of shadow-cast (defined above), r = distance in cm from filament to specimen, d = density of material in g/cc.

A convenient value for r is 15 cm, and for α is $\tan^{-1}\frac{1}{6}$. A thickness of 4A of uranium is then calculated to require about 12 mg of the element. It should be noted, of course, that any element of surface exposed at perpendicular incidence to the filament will receive, in this case, approximately 5 times as thick a deposit. This heavy deposit will distort the shapes of very small objects, by causing them to be somewhat elongated ellipsoids (if they are originally spheres) in the direction of the shadow-casting filament, and to become invisible beyond the "sunset edge" on the side away from the filament. A good estimate of the particle diameters, however, can be obtained by measurements perpendicular to the direction of shadowing.

c. The *evaporation* of chromium, manganese, gold, and uranium oxide is most easily performed from a 15–20 mil tungsten filament wound in the form of a conical helix, or basket, of three or four turns, and heated with about 20 amperes of current. The forming of the filament is most easily done by winding on a No. 4–6 woodscrew, with the tungsten and screw as hot as the hand will tolerate. The specimen can be mounted to receive the evaporating atoms either from the top or from the side of the conical filament.

Molten uranium and platinum wet and dissolve the filament, and a larger size wire, 30 mil, is desirable to prevent premature filament breakage. About 45 amperes of current are required to heat the filament adequately. A convenient filament form for the evaporation of these two elements is a sharp hairpin pointed directly toward the specimen, and with its apex bent up through 90°. The molten droplets will at first oscillate back and forth along the arms of the hairpin, but will finally come to rest and evaporate from the apex.

It is best to place the calculated amount of material in the filament and to evaporate it completely. The evaporating material should subtend as small a solid angle as possible at the specimen, in order to retain the sharpest possible shadow edge.

A vacuum satisfactory for vacuum aluminizing (about 10^{-4} mm mercury as measured on a Pirani gauge) is adequate for shadowcasting. The best criterion of excellence of vacuum, however, is the sharpness of the shadows, since too high a residual pressure will result in indistinct shadows. As indicated previously in this paper, Silicone oil should not be used in the diffusion pump, if pre-shadowed replicas are to be made, but Octoil is satisfactory.

d. The production of replicas of specimen mounted on glass: The technique for the stripping of collodion or Formvar replicas, with the aid of Scotch Tape, was described some years ago.[13] We have found this method very satisfactory, and can contribute little in addition. Collodion dissolved in amyl acetate, or Formvar dissolved in ethylene dichloride, can be applied either as a single drop on the pre-shadowed specimen held almost vertically, or the whole pre-shadowed glass slide can be immersed in the liquid and then allowed to dry while vertical. In either case, a satisfactory thickness of substrate film seems to be such that the film, while on the glass surface, appears to have a faintly purple hue.

The specimen screens for the final mounting of the replica are located on a strip of Scotch Tape, with a small piece of paper between the tape and the middle of the screen to avoid contact between the center of the screen and the tape. With wire-mesh screens the concave side should be toward the tape. The tape with specimen screen is held ready, and the moisture from one's breath is caused to condense on the collodion-coated glass slide. The condensate should be quite heavy, and there appears to be no danger of overdoing this. The tape with specimen is then quickly pressed down on the moisture-coated specimen, and pulled off fairly rapidly but gently. The successful removal of the pre-shadowing metal film can be detected prior to insertion of the specimen in the microscope by noting whether or not the stripped area of the glass slide appears more transparent than adjacent areas, since a film of gold or palladium-platinum only 5A thick has detectable optical opacity.

The specimen grid is removed from the Scotch Tape by gently running a sharp object, like a needle, around its periphery, and then removing the grid with flat-nosed tweezers. We have found that, for some obscure reason, the stripping of replicas from glass is most successful when done as soon as feasible after the shadow-casting and the application of the collodion film. No systematic difference in the technique of replica removal appears to exist between collodion and Formvar.

[13] V. J. Schaefer, Phys. Rev. **62**, 495 (1942).

The Determination of Stresses in Plastic Regions in Problems of Plane Flow*

P. S. SYMONDS
Division of Engineering, Brown University, Providence, Rhode Island
(Received August 5, 1948)

A simple numerical-graphical method for the calculation of stresses in plastic zones adjacent to boundaries of arbitrary shape is outlined, applicable to problems of plane plastic flow in ideally plastic materials. For illustration, the method is applied to determine the stresses in plastic zones near elliptic free boundaries. The yield load of a tensile bar containing elliptic notches is obtained for a series of notches of varying sharpness. Some related problems are discussed briefly.

INTRODUCTION

THIS paper presents an application of a simple numerical-graphical method for determining plastic stress distributions in cases of plane flow. As is well known, these stresses are statically determinate, and methods for their determination were given first by Hencky (1),** and Prandtl (2). Recently a number of new problems of plane flow have been treated, in particular by English and Russian investigators (see, for example, (3)–(5)).

In all but a few special cases of boundaries the stress solutions cannot be expressed in finite form, but must be obtained by some type of step-by-step integration, using the relations holding along the shear lines (characteristics). The numerical-graphical procedure to be described and applied in the following seems to be simpler and more rapid then any which has been previously described, and, as will be seen, is capable of high accuracy.

We shall be concerned here only with stress distributions in regions which are assumed to be wholly plastic; the extent of the plastic zone is not determined. Because of the hyperbolic character of the differential equations of plane plastic flow, the stresses in plastic zones near a boundary surface depend only on the shape of the boundary, and the stresses on the boundary. They are independent of the loading, i.e., of elastic stresses at large distances from the plastic zone considered. The loading and conditions on other boundary surfaces do affect the shape and size of the plastic zone. The complete solution requires the determination of the plastic-elastic interface; this is a difficult problem which has been successfully attacked in only a few cases (see, for example, (6)). In the present applications the plastic zone is assumed *a priori* to exist in certain regions. Displacements are not considered here; thus the solutions are those corresponding to the initial stages of flow adjacent to the given boundaries, before their shape has been significantly altered by the flow.

METHOD

In plane plastic flow we are concerned with stress components σ_x, σ_y, τ_{xy}, $\sigma_z=\frac{1}{2}(\sigma_x+\sigma_y)$, in rectangular axes x, y, z. The equations which the stresses must satisfy are two of equilibrium:

$$\partial\sigma_x/\partial x+\partial\tau_{xy}/\partial y=0; \quad \partial\sigma_y/\partial y+\partial\tau_{xy}/\partial x=0 \qquad [1]$$

and one expressing the yield condition:

$$(\sigma_x-\sigma_y)^2+4\tau_{xy}^2=4k^2, \qquad [2]$$

where k is the yield stress in pure shear. (In plane flow Eq. [2] expresses the yield condition corresponding to both the maximum shear stress and the maximum distortion energy theories.) It is convenient (8) to introduce new variables ω, θ by writing:

$$\begin{aligned}\sigma_x&=k(2\omega+\sin 2\theta),\\ \sigma_y&=k(2\omega-\sin 2\theta),\\ \tau_{xy}&=-k\cos 2\theta.\end{aligned} \qquad [3]$$

The yield condition is identically satisfied when this representation is used, and the equilibrium equations become

$$\begin{aligned}(\partial\omega/\partial x)+\cos 2\theta(\partial\theta/\partial x)+\sin 2\theta(\partial\theta/\partial y)&=0,\\ (\partial\omega/\partial y)+\sin 2\theta(\partial\theta/\partial x)-\cos 2\theta(\partial\theta/\partial y)&=0.\end{aligned} \qquad [4]$$

* This paper is based on a report prepared for Watertown Arsenal under a contract in Applied Mechanics.

** Numbers in parentheses refer to bibliography at end of paper.

The variables ω, θ have simple physical significance. $2k\omega$ is the mean of the normal stresses, while θ is the angle (positive counterclockwise) of the "first shear line" with the negative y axis (see Fig. 1); we here designate as "first shear line" the line whose normal makes a 45° angle (positive counterclockwise) with the direction of the maximum principal stress. The "second shear line" is perpendicular to the first, at a given point. As is well known, the system [4] possesses real characteristics, which are the two families of shear lines. If Eqs. [4] are transformed to orthogonal curvilinear coordinates in which s_1, s_2 represent distances along the first and second shear lines, respectively, they become

$$(d/ds_1)(\omega-\theta)=0, \quad (d/ds_2)(\omega+\theta)=0. \quad [5]$$

Thus along a first shear line the quantity $\omega-\theta$ is constant, while along a second shear line $\omega+\theta$ is constant.

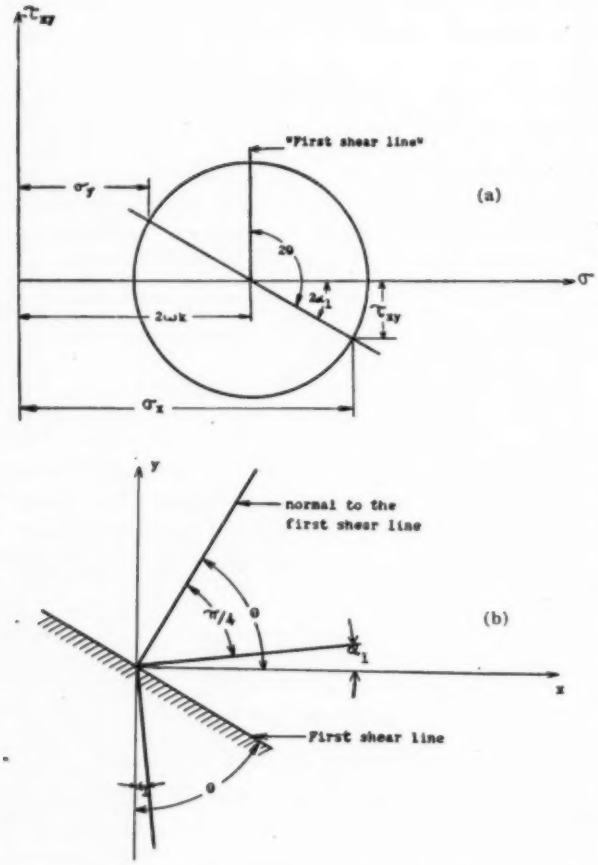


FIG. 1. (a) Mohr's circle diagram showing definitions of ω and θ; (b) direction of "first shear line."

At an unstressed boundary one of the principal stresses is zero, so that the shear lines meet the boundary curve at 45°. Moreover, $2\omega k=\pm k$, so that $\omega=\pm\frac{1}{2}$, the sign of ω depending on that of the non-zero principal stress. Other boundary conditions may be easily described.

Numerical Procedure

The numerical-graphical procedures are described with reference to an elliptical stress-free contour, for the sake of definiteness. We will take $\omega=+\frac{1}{2}$ on the ellipse, which means that the solutions we will obtain apply, for example, to the case of tensile loading perpendicular to the major axis of the ellipse. The method is easily applied to any given boundary contour, however. We define for convenience new variables ξ, η, as follows:

$$\xi\equiv\omega+\theta, \quad \eta\equiv\omega-\theta. \quad [6]$$

Thus the first and second shear lines are lines of constant η and ξ, respectively. The families of ξ, η lines near the end of the major axis of an elliptic contour are sketched in Fig. 2. By symmetry $\theta=3\pi/4$ for points on the x axis. A typical point in the plastic zone is specified by the ξ, η lines passing through the point. Thus $\widehat{QP}$ is the point at the intersection of the lines η_P, ξ_Q originating from points P, Q on the ellipse. The values of ω and θ at point $\widehat{QP}$ are easily determined from their known values at P and Q. Since $\xi_{\widehat{QP}}=\xi_P$ and $\eta_{\widehat{QP}}=\eta_P$, it follows that

$$\omega_{\widehat{QP}}=\tfrac{1}{2}+\tfrac{1}{2}(\theta_Q-\theta_P), \quad \theta_{\widehat{QP}}=\tfrac{1}{2}(\theta_Q+\theta_P). \quad [7]$$

Now if a series of points A, B, C, $\cdots$ are chosen on the ellipse, the net work of shear lines originating from them may be sketched, and the values ω, θ may be calculated using Eqs. [7] for all intersection points. It remains to determine the actual position of a point such as $\widehat{QP}$. For this a step-by-step graphical procedure is necessary, in lieu of an analytical expression for the shear lines. The following procedures have been suggested by Prager (7):

(a) A very simple but obviously rough procedure would be to locate a point $\widehat{QP}$ by drawing the lines tangent to the first shear line at P and to the second shear line at Q. Having located a row of intersection points in this manner one could then use the calculated values of ω and θ

for these points to locate the next row of points, and the process can then be repeated to cover as large an area as desired. A better approximation is easily obtainable, however.

(b) Instead of drawing a line tangent to the first shear line at P, one may average the angles of the tangents at P and at $\widehat{QP}$, and then draw from P a straight line at this angle, i.e., at $\bar{\theta}_P = \frac{1}{2}(\theta_P + \theta_{\widehat{QP}})$ with the negative y axis. Similarly, one may average the tangent angles at points Q and $\widehat{QP}$, and draw from Q a straight line at the angle $\bar{\phi}_Q = \frac{1}{2}(\theta_Q + \theta_{\widehat{QP}}) - \pi/2$ with the negative y axis. The process may then be repeated from the points already located, and the network extended as far as desired. If the shear lines were circular arcs, this method would locate the intersection points exactly.

(c) A third procedure is to calculate the radius of curvature of the shear line at each intersection point, and draw circular arcs having the proper radius, for example, from points P and Q, the intersection of which would locate $\widehat{QP}$. This method is probably capable of the same accuracy as method B, which seems preferable, however, on account of the smaller constructional difficulty involved.

Application

Method B was used in determining the shear line patterns for a series of ellipses with ratio a/b of major to minor axes equal to 1, 2, 4, and 6. In all these cases "base points" on the ellipses were taken arbitrarily at intervals of 5° in the ellipse angle β (the ellipse being defined by the equations $x = a\cos\beta$, $y = b\sin\beta$). In a few cases smaller intervals were used, but negligible differences were obtained by so doing. In the case of one ellipse, namely, that for which $a/b = 2$, both method B and the simpler method A were used, to obtain a comparison. In the case of the circle ($a/b = 1$) an analytic solution is available as a check on the numerical-graphical method. For the circular hole the shear lines are logarithmic spirals (by symmetry they intersect the radial lines at ±45°, and hence their equations are $dr = \pm r d\phi$, r, θ being polar coordinates). It is easily deduced that the normal stress $(\sigma_y)_0$ on the x axis is given by the expression

$$(\sigma_y)_0 = 2k(1 + \ln x/a). \qquad [8]$$

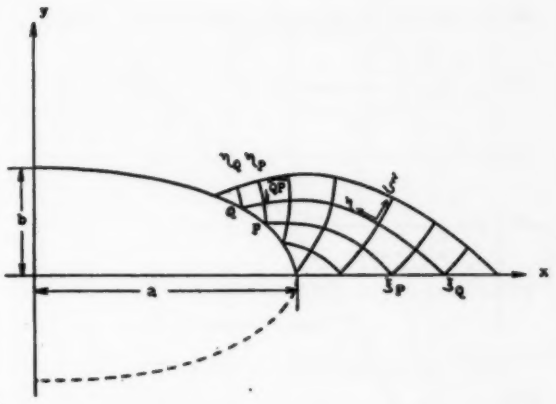


FIG. 2. Pattern of shear lines near an elliptic free boundary.

In constructing the shear line patterns by the numerical-graphical method, a "map-making" type of drafting machine was used, equipped with vernier graduated in minutes. From the results obtained, however, it would appear that an ordinary drafting machine with angle scale graduated in 5-minute or 30-minute intervals would be quite adequate for many purposes.

Results

Figure 3 shows the stresses $(\sigma_y)_0$ obtained for the ellipse $a/b = 2$ and the circle $a/b = 1$. The full curves for the ellipse $a/b = 2$ was obtained by method B and the dashed curve by the simplest method A. It is seen that the two methods yield stresses differing by only about 3 percent. This difference depends on the fineness of the mesh, however, and would be greater if points farther apart on the ellipse were used, say at intervals $\Delta\beta = 10°$. The curve for $a/b = 1$ in Fig. 3 was plotted from Eq. [8], and the points obtained by method B are shown for comparison. The differences in stress values are of the order of a few tenths of a percent, which indicates that method B is capable of very high accuracy even with a fairly coarse mesh.

Figure 4 shows the stress curves obtained for the series of ellipses $a/b = 1$, 2, 4, 6, and ∞. The straight line curve for the crack ($a/b = \infty$) is the value found by Hill (9),

$$\frac{1}{2k}(\sigma_y)_0 = 1 + \pi/2 \doteq 2.57. \qquad [9]$$

The stress curves of Fig. 4 may now be used to determine the yield load of a bar containing

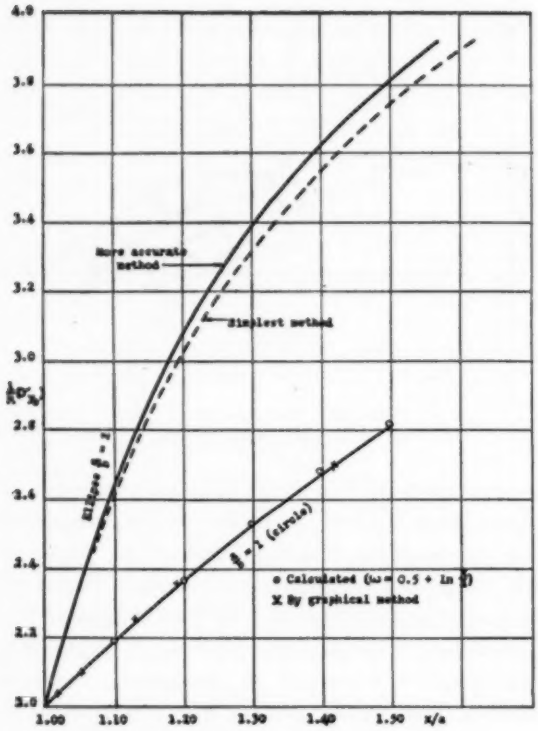


FIG. 3. Comparisons between stresses determined by two numerical-graphical methods, and by direct calculation (for case of a circular free boundary).

elliptic notches (placed symmetrically with respect to the bar axis). "Yield load" here means the force acting on the bar when the plastic zones from the notches have just spread to the axis of the bar. Figure 5 shows the yield load calculated for the particular case of a bar having half-width A equal to 1.5 times the notch depth a, plotted as a function of the ellipse ratio b/a. The load values were obtained by integrating graphically the plastic stress $(\sigma_y)_0$ over the width $2(A-a)$ of the bar at the notch section, using the curves of Fig. 4. According to Fig. 5 the yield load (in the sense used here) becomes larger as the notch is made sharper, although the loads required to *initiate* yield, of course, vary inversely with the notch sharpness.

LIMITATIONS ON RESULTS

A yield load calculated in the foregoing manner has significance only for a bar whose width $2A$ is less than a certain factor times the total notch depth $2a$, the value of the factor depending on the sharpness of the notches. The requirement is that

$$P=2\int_a^A (\sigma_y)_0 dx \leqq 4kA, \qquad [10]$$

where P is the load carried by the material at the notch section (just as the plastic zone has spread across the bar), and $4kA$ is the maximum load which can be carried by the bar at a cross section far removed from the notches. In the present problem A/a is limited as follows, for the two extreme cases of ellipse:

Circle $(b/a=1)$: $A/a \leqq e \doteq 2.72$;
crack $(b/a=0)$: $A/a \leqq 1+2/\pi \doteq 1.64$.

When A/a in a given problem exceeds the value set by Eq. [10], the stresses of Fig. 4 hold only within a certain distance from the edge of the notch. Presumably, in this case the plastic zone spreads across the bar in such a manner that a point on the axis off the notch section goes plastic before the point on the notch section. The precise location of the plastic regions would require, in general, the solution of the complete plastic-elastic problem. In a problem such as the present one relaxation methods present probably the only feasible means of attack.

It should be noted also that the stresses and yield loads obtained in the foregoing ignore the effect of the change in shape of the notch on the plastic stress distribution. This effect is probably unimportant when the notch is quite deep (A/a close to unity). Unfortunately, the determination of displacements for the general notch shape also

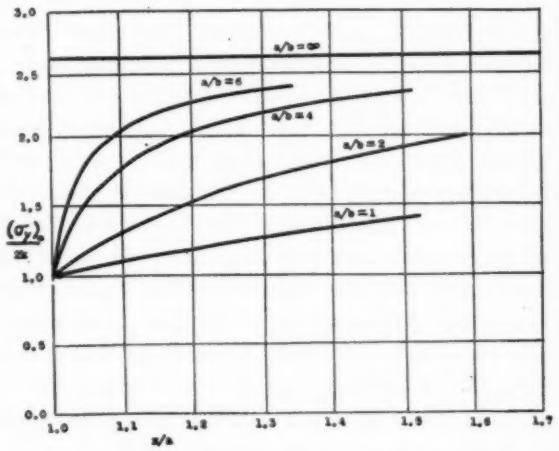


FIG. 4. Variation of stresses $(\sigma_y)_0$ over cross section of bar at notch, for ellipses having a/b = 1, 2, 4, 6, ∞.

depends on the prior solution of the plastic-elastic problem. As is well known, the differential equations defining the displacement rates are of hyperbolic type, and the displacement characteristics coincide with the shear lines (stress characteristics). However, the boundary values which determine the displacement rates in a given plastic region are those on the plastic-elastic interface, and hence are unknown until this surface is located. The special case of the infinitely sharp notch (crack) has been discussed by Hill (9).

RELATED PROBLEMS

The problem of determining the yield load of a tensile bar containing an elliptic hole may be considered in connection with the foregoing results. The yield loads plotted in Fig. 5 may be considered as a first approximation to those of a bar of width $2A=1.5(2a)$ containing an elliptic hole of diameter $2a$, and various degrees of flatness. However, the approximation in this case is worse than that in the external notch problem, since the plastic stress curves of Fig. 4 are incapable of satisfying the boundary conditions at the free straight edge of the bar, and hence are probably not even approximately correct near the free edge. Here a plastic region must be of the "constant state" type, in which the shear lines are straight and inclined at $\pm 45°$ with the straight edge of the bar. Between this zone and that governed by the interior elliptic free boundary there would probably be a third plastic zone in which the transition between the other regions occurs. Here again the error of the limit load obtained in a given case by using the stress curves of Fig. 4 can be determined only when the boundaries of the plastic zones are established.

Finally, it is of interest to consider the analogs in plane stress of the plane strain problems considered here, and in particular to enquire whether the results for plane strain can be easily adapted to plane stress. If we take the flow condition to be that of the maximum shear stress theory, then the equations of plane stress and of plane strain are identical as long as the principal stresses (in the plane stress case) are of opposite sign. Unfortunately, in the plane strain solutions obtained in the foregoing the

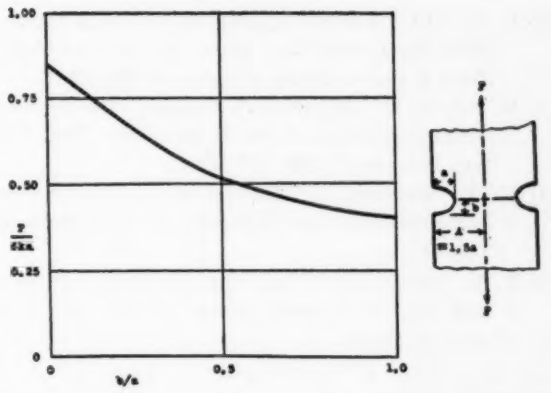


FIG. 5. Variation of yield load P with flatness of ellipse, for a notched bar of total notch depth $2a$, bar width $2A=1.5(2a)$. Yield load of main bar $=4kA=6ka$.

principal stresses are everywhere of the same sign, and hence do not yield plane stress solutions under this assumed flow condition. On the other hand, if we assume the maximum distortion energy as the flow condition, then the plane stress equations are different from those of plane strain. It is found that the plane stress equations analogous to Eqs. [1] and [2] are of hyperbolic type only when the principal stresses, σ_1, σ_2, obey the following relationship:

$$|\sigma_1-\sigma_2|>\tfrac{1}{3}|\sigma_1+\sigma_2|.$$

Thus the characteristics exist in a limited region near any free boundary. However, the families of characteristics are not orthogonal, and there does not seem to be any way to transform the equilibrium equations into the simple form of Eqs. [5]. Numerical integration, using the networks of characteristics, is still probably the best plan of attack on these problems, but seems to be considerably more laborious than in the case of plane strain. References (4) and (5) give examples of procedures and results.

ACKNOWLEDGMENT

The author is indebted to Miss Marion King, Research Assistant in the Graduate Division of Applied Mathematics at Brown University, for performing most of the computational work presented in this paper.

BIBLIOGRAPHY

(1) H. Hencky, "Ueber einige statisch bestimmte Faelle des Gleichgewichts in plastischen Koerpern," Zeits. f. angew. Math. u. Mech. 3, 241 (1923).

(2) L. Prandtl, "Anwendungsbeispiele zu einem Henckyschen Satz ueber das plastische Gleichgewicht," Zeits. f. angew. Math. u. Mech. **3**, 401 (1923).
(3) R. Hill, E. H. Lee, and S. J. Tupper, "The theory of wedge indentation of ductile materials," Proc. Roy. Soc. London (A) **188**, 273 (1947).
(4) V. V. Sokolovsky, "The theory of plasticity—an outline of work done in Russia," J. App. Mech. **13**, 1 (1946).
(5) V. V. Sokolovsky, "Theory of plasticity" (Russian with English chapter summaries), Izd. Ak. Nauk, Moscow (1946).
(6) L. A. Galin, "Plane elastic-plastic problems: plastic regions around circular holes in plates and beams," Prikl. Mat. i Mekh. **10**, 365 (1946) (in Russian).
(7) William Prager, "Plasticity for the aerodynamicist," J. Ae. Sci. **15**, 253 (1948).
(8) A. Nadai, *Plasticity—A Mechanics of the Plastic State of Matter* (McGraw-Hill Book Company, Inc., New York, 1931), Ch. 32.
(9) R. Hill, "The two-dimensional theory of plastic yielding of deeply-notched bars in tension," Armament Research Dept. Theor. Res. Rept. (September, 1946) (in press: to appear in Quarterly J. Mech. and App. Math.).

Erratum: Transmission Properties of Materials in the Millimeter Wave Region

[J. App. Phys. **19**, 1092 (1948)]

E. Wantuch, J. H. Rohrbaugh, J. P. Cooley, J. H. Greig, and J. Shmoys
New York University, New York, New York

November 19, 1948

THE authors have pointed out two errors in this paper:

(1) Figures 1 and 2 are interchanged.

(2) The caption for both figures should read "0.3 to 2.2 mm."

Heat Transfer in Sweat-Cooled Porous Metals*

SIDNEY WEINBAUM AND H. L. WHEELER, JR.
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received June 21, 1948)

Heat transfer inside sweat-cooled metals is analyzed and formulas showing the temperature distributions along the length of a sweat-cooled bar and across a sweat-cooled hollow cylinder are derived.

I. INTRODUCTION

AN answer to the ever-increasing demands for materials able to withstand very high temperatures may be sought either in the development of new materials or in the search for more efficient ways of cooling the heated materials. An important step toward the solution of this problem combines both of the above-mentioned methods and consists of the use of porous materials cooled by the flow of liquid or gas under pressure through the capillary passages inside the metal. This new development has come to be known as sweat cooling.[1]

The complete heat-transfer problem involved in sweat cooling is complex, and a better understanding of the whole problem may be achieved by isolating some of the constituent links for separate study. The investigation of heat transfer inside the sweat-cooled porous materials has been undertaken with this thought in mind.

II. HEAT TRANSFER IN A SWEAT-COOLED POROUS METAL BAR HEATED AT ONE END

In attacking the problem of heat transfer inside sweat-cooled porous materials, it is desirable to consider first the simplest case provided by a thin flat specimen of porous metal of thickness l (Fig. 1), one side of which is heated by a very hot flame, while cold fluid enters through the opposite side.

The heat supplied to the flat bar by the flame is conducted not only through the metal but also from the metal to the fluid flowing through the pores. The porous passages form a very complicated three-dimensional network, but for the purpose of this study an assumption is made of a simple network consisting of identical cylindrical channels running in parallel from one end of the specimen to the other.

The time rate of heat flow is, in the case of solid metal, proportional to the cross-sectional area perpendicular to the direction of flow and to the temperature gradient along the line of flow. For porous materials it is necessary to introduce another factor to take account of porosity. If porosity S is defined as the ratio of the volume of voids to the total volume of the porous metal specimen, then this factor is $1-S$ because, with the assumption made as to the character of the porous passages, the ratio of open area to total cross-sectional area is also equal to S. Thus the rate flow of heat may be represented by the equation

$$\begin{aligned} Q &= -\kappa(1-S)A(dT/dx) \\ &= -A_1(dT/dx) \end{aligned} \tag{1}$$

where κ is the thermal conductivity of the metal, A is the cross section of the specimen, and T is the temperature of the metal at any point in the specimen.

The rate of heat transfer from metal to cooling fluid is proportional to the area of contact

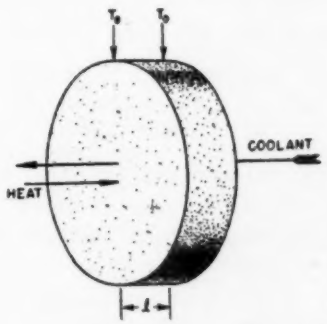


FIG. 1. Thin flat specimen of porous metal.

* This paper presents material drawn from Progress Reports No. 1–58 and 1–62 issued by the Jet Propulsion Laboratory.

[1] P. Duwez and H. E. Martens, "The powder metallurgy of porous metals and alloys having a controlled porosity," AIME Tech. Pub. No. 2343, 1948.

between metal and fluid and to the difference $T-t$ between the temperatures of metal and fluid. For steady heat flow the rates of heat flow along the metal and from metal to fluid must be the same at any particular point inside the porous metal. Since both T and t vary along the length of the metal specimen, this rate of heat transfer changes continuously along the length of the specimen. For an infinitesimally small length dx, the following expression holds:

$$\begin{aligned} dQ &= -h\pi NdA(T-t)dx \\ &= -A_2(T-t)dx \end{aligned} \tag{2}$$

where h is the coefficient of proportionality (or film coefficient), N is the number of passages per unit area, d is the diameter of the cylindrical passage, and πNd is the total circumference at any cross section.

The heat conducted from metal to fluid is used in raising the temperature of the fluid; hence

$$\begin{aligned} dQ &= Fc_p dt \\ &= A_3 dt \end{aligned} \tag{3}$$

where F is the total mass flow of fluid per unit time through the specimen, and c_p is the amount of heat necessary to raise the temperature of a unit mass of fluid 1 degree.

After differentiating Eq. (1) with respect to x and rewriting Eqs. (2) and (3) in a more convenient form, the following system of three simultaneous differential equations is obtained:

$$(dQ/dx) = -A_1(d^2T/dx^2) \tag{4}$$

$$(dQ/dx) = -A_2(T-t) \tag{5}$$

$$(dQ/dx) = A_3(dt/dx). \tag{6}$$

To solve for T, dQ/dx is eliminated from Eqs. (4) and (5):

$$A_1(d^2T/dx^2) - A_2(T-t) = 0. \tag{7}$$

Differentiating Eq. (7) gives

$$A_1\frac{d^3T}{dx^3} - A_2\frac{dT}{dx} + A_2\frac{dt}{dx} = 0. \tag{8}$$

Solving Eqs. (4) and (6) for dt/dx and substituting the obtained value of dt/dx in Eq. (8) result in a differential equation in T alone

$$\frac{d^3T}{dx^3} - b\frac{d^2T}{dx^2} - c\frac{dT}{dx} = 0 \tag{9}$$

where $b = A_2/A_3$ and $c = A_2/A_1$.

An equation for t is derived in a similar manner and is, as might have been expected, identical in form with Eq. (9)

$$d^3t/dx^3 - b(d^2t/dx^2) - c(dt/dx) = 0. \tag{10}$$

The solution of Eq. (9) is obtained by substituting $T = e^{\lambda x}$ for T and solving for λ

$$T = C_1 + C_2e^{\lambda_2 x} + C_3e^{\lambda_3 x} \tag{11}$$

where λ_2 and λ_3 are given by

$$\lambda = [b \pm (b^2+4c)^{\frac{1}{2}}]/2. \tag{12}$$

As only a link in the total heat-transfer problem is being examined, it is impossible to write down the boundary conditions. However, it is possible to perform an experiment in which the temperature of the incoming fluid t_l and the temperatures of the metal T_0 and T_l at both ends of the specimen are measured. By the use of these measured quantities, T and t inside porous metal can then be determined for these particular boundary conditions. Hence, for $x=0$,

$$C_1 + C_2 + C_3 = T_0 \tag{13}$$

and for $x=l$,

$$C_1 + C_2e^{\lambda_2 l} + C_3e^{\lambda_3 l} = T_l. \tag{14}$$

To make use of the third boundary condition, $t=t_l$ for $x=l$, Eqs. (4) and (5) can be solved for t in terms of T and d^2T/dx^2

$$t = -(A_1/A_2)(d^2T/dx^2) + T. \tag{15}$$

By differentiating Eq. (11) twice, an expression for d^2T/dx^2 can be obtained. After substitution of the expressions for T and d^2T/dx^2 in Eq. (15), this equation assumes the form, for $x=l$,

$$\begin{aligned} t_l = C_1 &+ C_2[1-(\lambda_2^2/c)]e^{\lambda_2 l} \\ &+ C_3[1-(\lambda_3^2/c)]e^{\lambda_3 l}. \end{aligned} \tag{16}$$

Solving Eqs. (13), (14), and (16) for C_1, C_2, and C_3 gives the following expressions for the constants of integration:

$$\begin{aligned} C_1 = \frac{1}{B}[&c(T_l-t_l)(e^{\lambda_3 l}-e^{\lambda_2 l}) \\ &-(T_l-T_0)(\lambda_3^2e^{\lambda_3 l}-\lambda_2^2e^{\lambda_2 l})] + T_0 \end{aligned} \tag{17}$$

$$C_2 = \frac{1}{B}[(T_l-T_0)\lambda_3^2e^{\lambda_3 l} - c(T_l-t_l)(e^{\lambda_3 l}-1)] \tag{18}$$

$$C_3 = \frac{1}{B}[c(T_l-t_l)(e^{\lambda_2 l}-1) - (T_l-T_0)\lambda_2^2e^{\lambda_2 l}] \tag{19}$$

where

$$B=\lambda_3^2e^{\lambda_3 l}(e^{\lambda_2 l}-1)-\lambda_2^2e^{\lambda_2 l}(e^{\lambda_3 l}-1). \quad (20)$$

It now becomes necessary to evaluate λ_2 and λ_3, and hence b and c:

$$b=(A_2/A_3)=(h\pi NdA/Fc_p), \quad (21)$$

$$c=(A_2/A_1)=[h\pi Nd/\kappa(1-S)]. \quad (22)$$

The numerical evaluation of quantities b and c for a sample experiment is given in Appendix A. The results are $b=1.071\times10^6$ and $c=5.74\times10^6$. Inserting the values of b and c in Eq. (12) gives the following numerical values for λ_2 and λ_3: $\lambda_2=1.071\times10^6$ and $\lambda_3=-5.36$.

Equations (17), (18), and (19) can be simplified by dividing numerators and denominators by the very large quantity $e^{\lambda_2 l}$ and neglecting the relatively very small additive terms in the resulting expressions. With $T_0=218°\text{F}$, $T_l=208°\text{F}$, and $t_l=68°\text{F}$, as given by the experiment, these expressions then become

$$C_1=\frac{-c(T_l-t_l)+(T_l-T_0)\lambda_2^2}{\lambda_3^2e^{\lambda_3 l}-\lambda_2^2(e^{\lambda_3 l}-1)}+T_0=122.8 \quad (23)$$

$$C_2=\frac{c(T_l-t_l)}{\lambda_2^2e^{\lambda_2 l}}=7.01\times10^{-4}e^{-22,310} \quad (24)$$

$$C_3=-C_1+T_0=95.2. \quad (25)$$

The fluid entering the specimen is at a much lower temperature than the cold end of the specimen. An assumption, that the entering fluid is of the same temperature as the cold wall, though it would simplify the problem greatly, seems to be in general quite unwarranted. The temperature of the fluid is, unquestionably, somewhat raised before entering the specimen; therefore the value of the initial temperature of the cooling fluid, $t=68°\text{F}$, may be regarded as that of the extreme case.

Equation (11) can be rewritten now as

$$T=122.8+7.01\times10^{-4}\exp(-22,310+1.071\times10^6x)+95.2\exp(-5.36x) \quad (26)$$

and utilizing Eq. (16) t becomes, similarly,

$$t=122.8-140\exp(-22,310+1.071\times10^6x)+95.2\exp(-5.36x) \quad (27)$$

where x is in feet. For the purposes of calculation it is convenient to put $x=y/48$ and rewrite Eqs. (26) and (27) as

$$T=122.8+7.01\times10^{-4}\exp[22,310(y-1)]+95.2\exp(-0.1117y) \quad (28)$$

and

$$t=122.8-140\exp[22,310(y-1)]+95.2\exp(-0.1117y) \quad (29)$$

where y varies from 0 to 1.

Examination of Eq. (28) reveals that the middle term does not contribute appreciably to the solution, its maximum value being 7.01×10^{-4} for $y=0$. Thus

$$T=122.8+95.2\exp(-0.1117y). \quad (30)$$

This middle term, however, becomes very important in Eq. (29). For all values of y, except those exceedingly close to $y=1$, this term is negligible, but for $y\to1$ it begins to contribute and reaches the value of 140 for $y=1$. Thus it becomes apparent that the temperatures of metal and fluid are indistinguishable throughout most of the thickness of the metal, and that the heating of the cooling fluid occurs in the first very small fraction of the volume of the metal at the end where the fluid enters the porous material. In general, the character of the solution is such that the increase in flow would not change the general physical picture as represented by equations for the temperatures of metal and fluid. As can be seen in Appendix A, constants b and c vary only as fractional powers of flow, whereas even a considerable change in the values of b and c would affect the character of the solution but little.

Equations (28) and (29) differ only in their middle terms, and thus a calculation can be easily made of the distance the fluid travels before the difference in temperature between metal and fluid becomes, say, 0.01°F:

$$140\exp[22,310(y-1)]=0.01 \quad (31)$$

or

$$1-y=\frac{\ln14,000}{22,310}=0.000425. \quad (32)$$

This is equivalent to about 0.0001 inch and corresponds to an area of about 1 sq. in. of contact between metal and fluid.

Since the temperatures T and t are indistinguishable except at the cool end of the specimen,

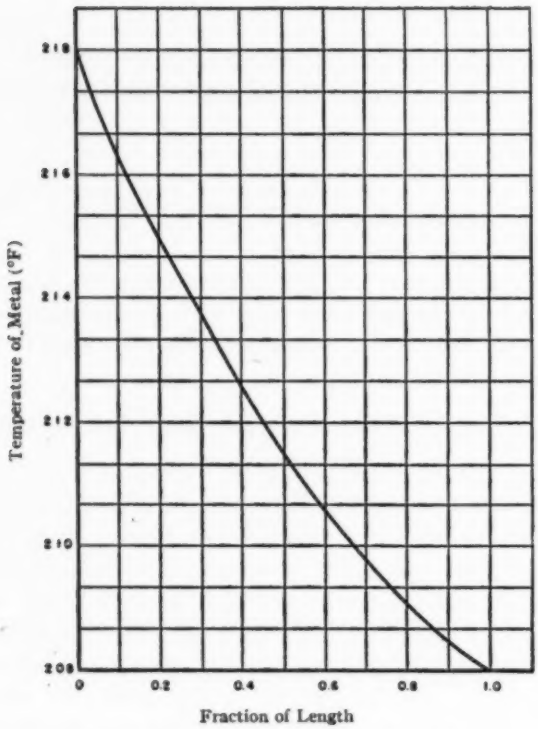


FIG. 2. Temperature of metal along the length of a sweat-cooled bar.

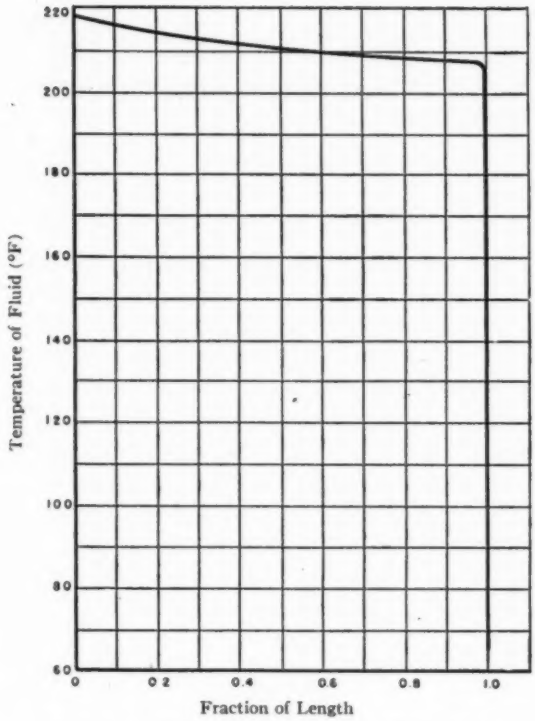


FIG. 3. Temperature of cooling fluid along the length of a sweat-cooled bar.

they are plotted separately. Figure 2 shows the temperature of metal, and Fig. 3 on a smaller temperature scale shows the temperature of fluid as a function of the fractional distance along the axis of the test specimen.

III. HEAT TRANSFER IN A SWEAT-COOLED POROUS METAL HOLLOW CYLINDER

Specimens of actual interest in sweat-cooling studies are hollow cylinders cooled by radial flow of fluid. The flame travels lengthwise through the hollow cylinder (Fig. 4), and the metal is prevented from melting by fluid flowing under pressure radially through the pores, from the cooler outer cylindrical surface to the hotter inner surface. Three differential equations analogous to those for the case of a bar heated at one end can be written

$$\frac{dQ}{dr}=-B_1\frac{dT}{dr}-B_1r\frac{d^2T}{dr^2} \tag{33}$$

$$\frac{dQ}{dr}=-B_2r(T-t) \tag{34}$$

$$\frac{dQ}{dr}=B_3\frac{dt}{dr} \tag{35}$$

where

$$B_1=2\pi kl(1-S) \tag{36}$$

$$B_2=2\pi^2Ndlh \tag{37}$$

$$B_3=Fc_p. \tag{38}$$

The constant B_2 in Eq. (36) contains the film coefficient h, which for the purpose of this calculation is considered constant, as it varies but slowly with r (see Appendix B). It is possible, however, as shown in Appendix C, to include in Eq. (34) the dependence of h on the radius and still solve the resulting system of equations.

To solve the above system of equations for T, Eqs. (33) and (34) are first solved for t:

$$t=T-\frac{B_1}{B_2}\frac{1}{r}\frac{dT}{dr}-\frac{B_1}{B_2}\frac{d^2T}{dr^2}. \tag{39}$$

From Eqs. (34) and (35)

$$(dt/dr)=-(B_2/B_3)r(T-t) \tag{40}$$

and differentiating Eq. (39) with respect to r

results in

$$\frac{dt}{dr}=\frac{dT}{dr}+\frac{B_1}{B_2}\frac{1}{r^2}\frac{dT}{dr}-\frac{B_1}{B_2}\frac{1}{r}\frac{d^2T}{dr^2}-\frac{B_1}{B_2}\frac{d^3T}{dr^3}. \quad (41)$$

Substituting in Eq. (40) the expressions for t and dt/dr from Eqs. (39) and (41), respectively, and performing the necessary algebraic manipulations give a differential equation in T only

$$r^2\frac{d^3T}{dr^3}+(-br^3+r)\frac{d^2T}{dr^2}-[(b+c)r^2+1]\frac{dT}{dr}=0 \quad (42)$$

where $b=B_2/B_3$ and $c=B_2/B_1$.

By eliminating T and its derivatives from Eqs. (33), (34), and (35), an equation similar to Eq. (42) can be obtained for the temperature of the cooling fluid

$$r^2\frac{d^3t}{dr^3}-(br^3+r)\frac{d^2t}{dr^2}+[-(b+c)r^2+1]\frac{dt}{dr}=0. \quad (43)$$

Equations (42) and (43) are of the general form

$$x^2\frac{d^3y}{dx^3}+(Ax^3+Bx)\frac{d^2y}{dx^2}+(Cx^2-B)\frac{dy}{dx}=0 \quad (44)$$

which is reducible to Kummer's first confluent hypergeometric equation[2]

$$x\frac{d^2y}{dx^2}+(\alpha-x)\frac{dy}{dx}-\beta y=0 \quad (45)$$

by the following procedure.

It may be ascertained by inspection that a power-series solution of Eq. (44) would consist of terms with exponents of x increasing in steps of 2. An even-power solution of this equation may be expressed in the form

$$y=F(mx^2)=F(u) \quad (46)$$

where m is a constant. Then

$$\frac{dy}{dx}=2mx\frac{dF}{du} \quad (47)$$

$$\frac{d^2y}{dx^2}=2m\frac{dF}{du}+4m^2x^2\frac{d^2F}{du^2} \quad (48)$$

$$\frac{d^3y}{dx^3}=12m^2x\frac{d^2F}{du^2}+8m^3x^3\frac{d^3F}{du^3}, \quad (49)$$

and Eq. (44) becomes

$$u^2\frac{d^3F}{du^3}+\left(\frac{A}{2m}u^2+\frac{B+3}{2}u\right)\frac{d^2F}{du^2}+\frac{A+C}{4m}u\frac{dF}{du}=0. \quad (50)$$

The constant m is arbitrary; hence, by putting $m=-A/2$ and dividing by u, Eq. (50) is reduced to

$$u\frac{d^3F}{du^3}+\left(-u+\frac{B+3}{2}\right)\frac{d^2F}{du^2}-\frac{A+C}{2A}\frac{dF}{du}=0 \quad (51)$$

which is equivalent in form to Eq. (45).

The solution of the confluent hypergeometric Eq. (45) is usually written[3] as ${}_1F_1(\beta;\alpha;x)$ where

$${}_1F_1(\beta;\alpha;x)=1+\frac{\beta}{1!\alpha}x+\frac{\beta(\beta+1)}{2!\alpha(\alpha+1)}x^2+\cdots. \quad (52)$$

Hence the solution of Eq. (51) is

$$\frac{dF}{du}={}_1F_1\left(\frac{C}{2A}+\frac{1}{2};\frac{B}{2}+\frac{3}{2};u\right). \quad (53)$$

But in general[4]

$$\frac{d}{du}{}_1F_1(\beta;\alpha;u)=\frac{\beta}{\alpha}{}_1F_1(\beta+1;\alpha+1;u), \quad (54)$$

and hence

$${}_1F_1\left(\frac{C}{2A}+\frac{1}{2};\frac{B}{2}+\frac{3}{2};u\right)=\frac{A(B+1)}{C-A}{}_1F_1\left(\frac{C}{2A}-\frac{1}{2};\frac{B}{2}+\frac{1}{2};u\right). \quad (55)$$

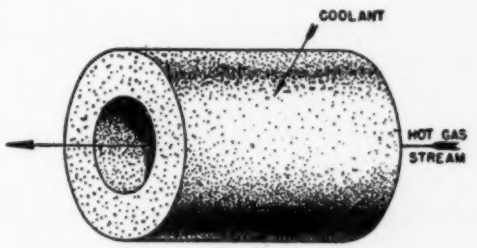


FIG. 4. Hollow cylindrical specimen.

[2] E. G. C. Poole, *Introduction of the Theory of Linear Differential Equations* (Clarendon Press, Oxford, 1936), p. 139.

[3] Reference 2, p. 140.

[4] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (University Press, Cambridge, 1920), p. 275.

Then Eq. (46) takes the form

$$y={}_1F_1\left(\frac{C}{2A}-\frac{1}{2};\frac{B}{2}+\frac{1}{2};\frac{-A}{2}x^2\right). \quad (56)$$

For $B=-1$, which is the value B has in Eq. (43), the solution represented by Eq. (56) breaks down. For Eq. (42), where $A=-b$, $B=1$, and $C=-(b+c)$, Eq. (56) becomes

$$T={}_1F_1\left(\frac{c}{2b};1;\frac{b}{2}r^2\right). \quad (57)$$

The constants b and c, which depend on the exceedingly large inside surface of the porous metal, turn out to be very large quantities themselves, as shown in a sample calculation in Appendix B; thus the series represented by Eq. (57) is divergent. The second solution of Eq. (45) could be derived from the known solution (Eq. (57)) and would also be divergent. However, asymptotic expansions for the confluent hypergeometric function are known[5] to be

$$F_1\sim\frac{e^u}{u^{\alpha-\beta}}{}_2F_0\left(\alpha-\beta;1-\beta;\frac{1}{u}\right) \quad (58)$$

and

$$F_2\sim\frac{1}{u^{\beta}}{}_2F_0\left(\beta;1+\beta-\alpha;-\frac{1}{u}\right) \quad (59)$$

where

$${}_2F_0(\beta;\alpha;u)=1+\frac{\beta\alpha}{1!}u+\frac{\beta(\beta+1)\alpha(\alpha+1)}{2!}u^2+\cdots. \quad (60)$$

Thus two of the solutions of Eq. (42) can be written as

$$T_1=e^{br^2/2}r^{-2+(c/b)}{}_2F_0\left(\frac{c}{2b};1-\frac{c}{2b};\frac{2}{br^2}\right) \quad (61)$$

and

$$T_2=r^{-(c/b)}{}_2F_0\left(\frac{c}{2b};\frac{c}{2b};-\frac{2}{br^2}\right). \quad (62)$$

[5] E. Jahnke and F. Emde, *Tables of Functions* (Dover Publications, New York, 1943), p. 275.

The third solution is

$$T_3=\text{constant} \quad (63)$$

as Eq. (42) contains no term with T in it.

For very large values of b, it is necessary to take only the first term in the two series occurring in Eqs. (61) and (62). As the first term of a hypergeometric series is equal to 1, the complete solution of Eq. (42) takes the form

$$T=C_1e^{(b/2)r^2}r^{(c/b)-2}+C_2r^{-(c/b)}+C_3. \quad (64)$$

To determine the physical character of this solution it is again necessary to utilize the experimental values for T_i and T_0, the metal temperatures on the inside and outside surfaces of the hollow cylinder, respectively, and t_0, the temperature of the cooling fluid at the surface of entrance. From Eqs. (33) and (34), an equation for t in terms of T and its derivatives can be obtained; and this equation in turn, by utilizing the expression for T as given by Eq. (64) and its derivatives, can be written as

$$t=C_1e^{br^2/2}r^{c/b}\left[-\frac{b^2}{c}+\left(-1+\frac{2b}{c}\right)\frac{1}{r^2}-\frac{1}{c}\left(\frac{c}{b}-2\right)^2\frac{1}{r^4}\right]+C_2r^{-(c/b)}\left(1-\frac{c}{b^2r^2}\right)+C_3. \quad (65)$$

The first terms in both sets of brackets in Eq. (65) are very much larger, for reasonable values of r, than the rest of the additive terms. Neglecting these last terms, this equation can be written as

$$t=-C_1\frac{b^2}{c}e^{br^2/2}r^{c/b}+C_2r^{-(c/b)}+C_3. \quad (66)$$

The use of boundary conditions in Eqs. (64) and (66) gives

$$T_1=C_1e^{br_1^2/2}r_1^{(c/b)-2}+C_2r_1^{-(c/b)}+C_3 \quad (67)$$

$$T_0=C_1e^{br_2^2/2}r_2^{(c/b)-2}+C_2r_2^{-(c/b)}+C_3 \quad (68)$$

$$t_0=-C_1\frac{b^2}{c}e^{br_2^2/2}r_2^{c/b}+C_2r_2^{-(c/b)}+C_3 \quad (69)$$

where r_1 and r_2 are, respectively, the inside and outside radii of the hollow cylinder. Equations (67), (68), and (69) yield the following expres-

sions for the constants of integration:

$$C_1 \sim \frac{c}{b^2}(T_0 - t_0)e^{-(br_2^2/2)}r_2^{-(c/b)} \quad (70)$$

$$C_2 \sim \frac{T_i - T_0}{r_1^{-(c/b)} - r_2^{-(c/b)}} \quad (71)$$

$$C_3 \sim T_0 - \frac{(T_i - T_0)r_2^{-(c/b)}}{r_1^{-(c/b)} - r_2^{-(c/b)}}. \quad (72)$$

The values of r_1 and r_2, for a specimen used in a series of sweat-cooling experiments, were $r_1 = 0.0417$ and $r_2 = 0.0742$ foot, respectively. The calculations of constants b and c, in terms of flow of the cooling fluid, are given in Appendix B; these calculations* result in values $b = 1.152 \times 10^8 F^{-1}$ and $c = 1.326 \times 10^4 F^{\frac{1}{2}}$, where F is in cubic centimeters per minute.

For a typical run of $F = 20$ cc/min., the constants become $b = 1.564 \times 10^7$ and $c = 3.60 \times 10^4$. The observed temperatures for this run were $T_i = 210°F$, $T_0 = 170°F$, and $t_0 = 70°F$. Substituting these numerical values in Eqs. (70), (71), and (72) results in the following values for the constants of integration

$$C_1 = 1.48 \times 10^{-8} \exp(-4.30 \times 10^4)$$
$$C_2 = 30{,}100$$
$$C_3 = 170 - 30{,}100 \times 0.0742^{-2.30 \times 10^{-3}}$$

and Eqs. (64) and (66) can be rewritten as

$$T = 1.48 \times 10^{-8} \exp\left(4.30 \times 10^4 \times \left[\left(\frac{r}{0.0742}\right)^2 - 1\right]\right) r^{-1.998} + 3.026 \times 10^4 \times \left[\left(\frac{r}{0.0742}\right)^{-2.30 \times 10^{-3}} - 1\right] + 170, \quad (73)$$

and

$$t = -100\left(\frac{r}{0.0742}\right)^{2.30 \times 10^{-3}} \exp\left(4.30 \times 10^4 \times \left[\left(\frac{r}{0.0742}\right)^2 - 1\right]\right) + 3.026 \times 10^4 \times \left[\left(\frac{r}{0.0742}\right)^{-2.30 \times 10^{-3}} - 1\right] + 170. \quad (74)$$

* The calculation of constants b and c depends on a number of approximations, but even a large variation in the numerical values of these constants has very little effect on final results.

The first term in Eq. (73) is very small and can be neglected; hence

$$T = 3.026 \times 10^4 \times \left[\left(\frac{r}{0.0742}\right)^{-2.30 \times 10^{-3}} - 1\right] + 170. \quad (75)$$

For calculation purposes it is convenient to write

$$r/0.0742 = 1 - y \quad (76)$$

and expand T as a series in y. The result is

$$T = 69.6 \sum_{q=1}^{\infty} \frac{y^q}{q} + 170. \quad (77)$$

In the case of the specimen under consideration, only three terms of the summation are needed for an accuracy of 1°F.

By subtracting Eq. (74) from Eq. (75), an expression for the difference in temperatures of metal and cooling fluid is obtained:

$$T - t = 100\left(\frac{r}{0.0742}\right)^{2.30 \times 10^{-3}} \times \exp\left(4.30 \times 10^4 \left[\left(\frac{r}{0.0742}\right)^2 - 1\right]\right). \quad (78)$$

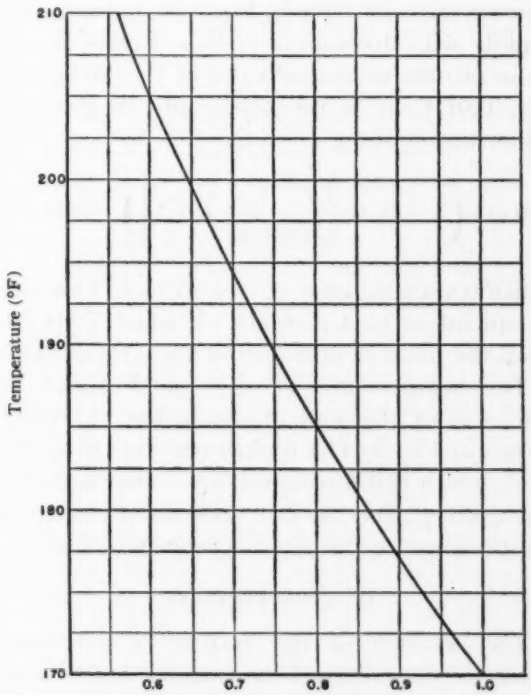


FIG. 5. Temperature of metal inside a sweat-cooled hollow porous metal cylinder.

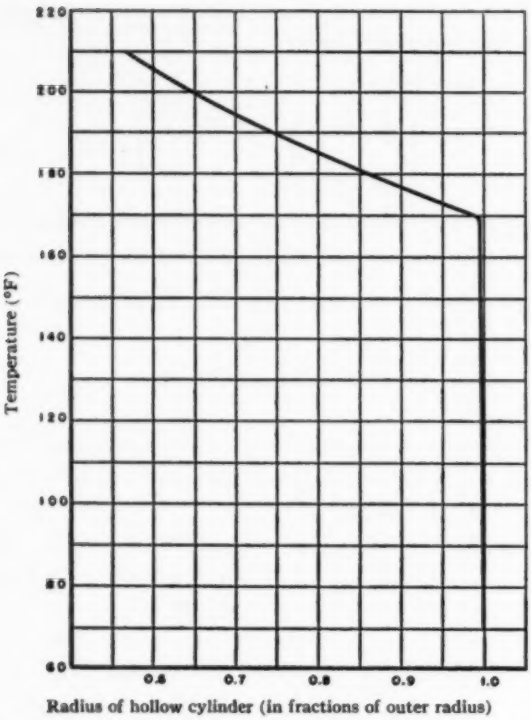


FIG. 6. Temperature of cooling fluid inside a sweat-cooled hollow porous metal cylinder.

The exponential term in Eq. (78) decreases very rapidly with decrease in r. To calculate at what value of r the numerical value of $T-t$ reduces to, say, 0.01°F, it is necessary only to solve the following equation

$$100\exp\left(4.30\times10^{4}\left[\left(\frac{r}{0.0742}\right)^{2}-1\right]\right)=0.01 \quad (79)$$

which is satisfied for $r=0.074192$ foot. This value is equivalent to a distance of about 0.001 inch from the point of entrance of the cooling fluid.

The temperature of the metal across the thickness of the wall of the hollow cylinder is shown in Fig. 5. The temperature of the cooling fluid, which is indistinguishable from that of the metal except for the last very small fraction of the thickness of the metal, is shown in Fig. 6.

IV. CONCLUSIONS

The solution of the system of differential equations describing heat transfer inside a bar during sweat cooling indicates that, for the volume of flow that may be expected during sweat cooling, the temperatures of metal and fluid are indistinguishable throughout practically the entire volume of porous metal. The temperature distribution along the length of the bar is shown to be not linear, as in the case of a solid metal bar, but dependent upon an exponential term.

The physical picture of heat transfer between metal and cooling fluid in a hollow porous metal cylinder sweat cooled by a radial flow is the same as in the case of linear flow of sweat-cooling fluid through a porous metal bar. The temperature distribution in the radial direction in the metal depends not upon a logarithmic term, as in the case of solid metal, but upon a fractional power of the radius.

APPENDIX A

Calculation of Constants b and c for a Bar

Constants b and c are given by Eqs. (21) and (22)

$$b=\frac{h\pi NdA}{Fc_p}, \quad c=\frac{h\pi Nd}{\kappa(1-S)}.$$

N, d, and S in the above expressions are properties connected with the porosity of the metal. From the study of flow of gases through porous steel of porosity $S=0.403$, it was found that for low gas pressures the flow obeys Poiseuille's equation for streamline flow. For a single cylindrical channel the streamline flow is[6]

$$\frac{dn}{dt}=\frac{\pi d^{4}(p_1^{2}-p_2^{2})}{256l\eta RT} \quad (A\text{-}1)$$

where dn/dt is the flow (in lb. moles/sec.), η is the coefficient of viscosity of the gas, l is the thickness of the specimen (in inches), and p_1 and p_2 are the pressures (in lb./sq. in.) at the two ends of the specimen. For hydrogen, $\eta=1.29$ lb. sec./sq. in. at 20.8°C; the thickness of the specimen used is $l=0.25$ inch; and $(dn/dt)/(p_1^{2}-p_2^{2})$ as obtained from the experimental data on flow is equal to 2.07×10^{-7}. Assuming a network of porous passages which consists of identical cylindrical channels and solving Eq.

[6] R. M. Barrer, *Diffusion in and through Solids* (University Press, Cambridge, 1941).

(A-1) for d, the following expression is obtained:

$$d=\left[\frac{64l\eta RT(dn/dt)}{S(p_1^2-p_2^2)}\right]^{\frac{1}{4}}=2.40\times10^{-4} \quad \text{(A-2)}$$

where porosity S is equal to the number of holes N per square inch multiplied by the area of a single hole

$$S=N\pi d^2/4. \quad \text{(A-3)}$$

From Eq. (A-3)

$$N=4S/\pi d^2=8.92\times10^6. \quad \text{(A-4)}$$

The experimental data used in the report were taken with a flat cylindrical steel specimen and water as cooling fluid; thus $c_p=1$, and the coefficient of heat conductivity κ is taken as $\kappa=16$ B.t.u./hr. ft.°F for the particular steel used. The thickness and the diameter of the specimen were $l=0.25$ inch and $D=1\frac{3}{8}$ inch, respectively.

There is no very satisfactory way of calculating h. Neglecting the change in viscosity with temperature,[7] the formula used is

$$hd/k=1.86\left(\frac{d^2Gc_p}{kl}\right)^{\frac{1}{3}} \quad \text{(A-5)}$$

where h is the coefficient of heat transfer from metal to fluid (in B.t.u./hr. sq. ft.°F); k is the thermal conductivity of the fluid (in B.t.u./hr. ft.°F); G is the mass velocity (in lb./hr. sq. ft.); c_p is specific heat at constant pressure (in B.t.u./lb. fluid °F); d is the inside diameter of the tube (in feet); and l is the heated length of the tube (in feet).

G may be conveniently expressed in terms of total flow F_1, which has been measured in cubic centimeters per minute:

$$G=\frac{F_1\times2.20\times10^{-3}\times60\times4\times4\times144}{N\pi D^2\pi d^2}$$

$$=31.5F_1 \text{ lb./hr. sq. ft.} \quad \text{(A-6)}$$

Solving Eq. (A-5) for h gives, with $k=0.398$ for water,

$$h=1.86\left(\frac{k^2Gc_p}{dl}\right)^{\frac{1}{3}}=426F_1^{\frac{1}{3}}. \quad \text{(A-7)}$$

[7] W. H. McAdams, *Heat Transmission* (McGraw-Hill Book Company, Inc., New York, 1942), p. 190.

Thus b and c may now be calculated as functions of flow

$$b=2.70\times10^6F_1^{-\frac{1}{3}}, \quad \text{(A-8)}$$

and

$$c=3.62\times10^6F_1^{\frac{1}{3}}. \quad \text{(A-9)}$$

A number of sweat-cooling runs have been made with the above specimen. A representative run of 4 cc/min. has been selected for a sample calculation. The temperature readings at the hot and cold ends of the specimen for this run were 218 and 208°F, respectively. Substituting $F=4$ in Eqs. (A-8) and (A-9) results in

$$b=1.07\times10^6 \quad \text{and} \quad c=5.74\times10^6.$$

APPENDIX B

Calculation of Constants b and c for a Hollow Cylinder

The constants b and c are $b=B_2/B_3$ and $c=B_2/B_1$ where B_1, B_2, and B_3 are defined by Eqs. (36), (37), and (38)

$$B_1=2\pi\kappa l(1-S), \quad B_2=\frac{8\pi Slh}{d}, \quad B_3=Fc_p.$$

The porous specimen used in this series of sweat-cooling experiments is a hollow copper cylinder of length $l=1.67$ inches and porosity $S=0.403$; the thermal conductivity of copper is taken as $\kappa=220$ B.t.u./hr. ft.°F. The film coefficient h is calculated from Eq. (A-5). The mass velocity per tube, on the basis of the definition used for the average pore, is variable across the thickness of the cylinder; hence it is necessary to use some average value, say, the value at the distance $(r_1+r_2)/2$ from the center, where r_1 and r_2 are the inside and the outside radii of the cylinder, respectively. In terms of total flow F_1, as measured in cubic centimeters per minute, this mass velocity is equal to $G=6.47F_1$.

Substituting this value of G in Eq. (A-5) gives, in the case where the fluid is water, $h=216F_1^{\frac{1}{3}}$.

Expressions for B_1, B_2, and B_3 become now $B_1=1148$, $B_2=1.522\times10^7F_1^{\frac{1}{3}}$, and $B_3=1.320\times10^{-1}F_1$. Hence $b=1.152\times10^8F_1^{-\frac{1}{3}}$ and $c=1.326\times10^4F_1^{\frac{1}{3}}$.

APPENDIX C

Solution for the Case Where h Is Taken as Dependent on r

There is no very satisfactory way of calculating the film coefficient h, and the different formulas for h vary as to the power of G and hence of r. Thus, in introducing this dependence of h upon r into Eq. (37) it is more satisfactory to use the relationship

$$h \sim r^n \tag{C-1}$$

where n may assume any desired value.

Film coefficient appears only in Eq. (34), which now takes the form

$$dQ = -B_2'(T-t)r^{n+1} \tag{C-2}$$

where B_2' is a coefficient different from A_2.

Equation (C-2) takes the place of Eq. (34) in the system of three differential equations (33), (34), and (35). By the same procedure that was used in obtaining Eq. (42), the system of Eqs. (33), (C-2), and (35) is reduced to

$$r^2\frac{d^3T}{dr^3} - [br^{n+3} + (n-1)r]\frac{d^2T}{dr^2} - [(b+c)r^{n+2} + (n+1)]\frac{dT}{dr} = 0. \tag{C-3}$$

By inspection, a solution of the following type may be assumed:

$$T = F(mr^{n+2}) = F(v). \tag{C-4}$$

After differentiating Eq. (C-4) three times with respect to r and substituting the obtained expression for the successive derivatives of T in Eq. (C-3), the latter becomes

$$v\frac{d^3F}{dv^3} + \left[-\frac{b}{m(n+2)}v + 2\right]\frac{d^2F}{dv^2} - \frac{b(n+2)+c}{m(n+2)^2}\frac{dF}{dv} = 0. \tag{C-5}$$

For $m = b/(n+2)$, Eq. (C-5) becomes

$$v\frac{d^3F}{dv^3} + (-v+2)\frac{d^2F}{dv^2} - \frac{b(n+2)+c}{b(n+2)}\frac{dF}{dv} = 0 \tag{C-6}$$

which is of the same type as Eq. (51) and reduces to Eq. (51) for $n = 0$.

Equation (C-6) is satisfied by the confluent hypergeometric series

$$T = {}_1F_1\left[\frac{c}{b(n+2)}; 1; \frac{br^{n+2}}{n+2}\right]. \tag{C-7}$$

For $n = -1/3$, which is the value of n in Appendix B,

$$T = {}_1F_1\left(\frac{3c}{5b}; 1; \frac{3br^{5/3}}{5}\right). \tag{C-8}$$

Letters to the Editor

A New Method of Capacity Measurement on Dry Disk Rectifiers

K. Lehovec
Signal Corps Engineering Laboratories, Fort Monmouth, New Jersey
October 25, 1948

WITH an a.c. voltage (and a suitable d.c. bias) applied to a dry disk rectifier in series with a small resistor, a part of the rectifier characteristic is registered on a cathode-ray oscillograph using the voltage at the rectifier as abscissa and the voltage at the resistor as ordinate. The characteristic shows a "loop" (Fig. 1) which is due to the capacity of the rectifier. From this loop the capacity (and the resistance) of the rectifier can be calculated as a function of the voltage assuming that

a. the rectifier is equivalent to a capacity shunted by a resistance, and that

b. both are functions of the momentary voltage only. If the applied voltage has a sine-shape in the considered range, we have

$$C=\Delta I/(2w(v_0^2-v^2)^{\frac{1}{2}}).$$

(ΔI width of the loop at the voltage v, w frequency, v_0 peak of the a.c. voltage; the capacity C is here defined as the small change in charge divided by the corresponding small change in voltage.) By this method slow changes in the barrier layer taking place in more than $2\pi/w$ sec. are subpressed. Comparison with measurements by the usual bridge method[1] shows that the change of the capacity of selenium rectifiers with voltage is much larger in the "static" bridge measurements than in the "dynamic" loop measurements. Thus we conclude that the capacity is determined partly by effects with very short time constant ($\ll$1/1000 sec.) and partly by effects with considerable time constant ($\gtrsim$1/1000 sec.), the latter ones may add as much as 30 percent to the capacity at zero voltage. Additional proof of time lag effects with long time constants is given by the temporary disturbance of a balanced bridge after a voltage pulse has been sent through a selenium rectifier located in one branch of the bridge.

The time lag effects dealt with in this paper are considered to be effected by trapping centers[2] rather than by dislocation of ions[3] in the barrier layer.

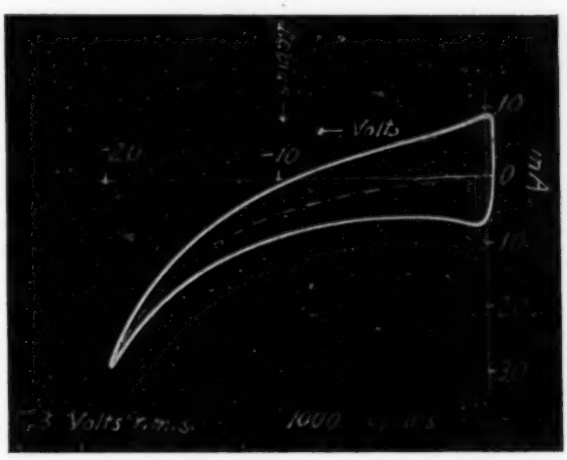


FIG. 1.

[1] Schottky and Deutschmann, Physik. Zeits. 30, 893 (1929).
[2] A. W. Lawson, *Ionization of Donator Levels in Crystal Rectifiers by Thermal Agitation*, NDRC 14-173, University of Penna., July 7, 1943.
[3] F. Rose and H. Schmidt, Zeits. f. Naturforschung **2a**, 226 (1947).

"A Preliminary Study of a Physical Basis of Bird Navigation"

William C. Schumacher
Office of Naval Research, New York, New York
October 14, 1948

THE original article[1] and the discussion by Messrs. Slepian, Varian, and Davis (as found in the March 1948 issue) have just come to my attention, inasmuch as I am, unfortunately, not a steady reader of or subscriber to your excellent publication.

My purpose in writing is to point out a possibility that appears to have escaped notice in the aforementioned references, and (unless this possibility has been mentioned by others in the intervening time) the present communication might initiate an investigation by someone better prepared to pursue such an investigation than I am.

Each of the writers discussed the movement of the bird through the earth's magnetic field as if the bird were a rigid object, with the obvious result that only the well-known induced (static) e.m.f. received consideration. (Davis, to be sure, suggested that an e.m.f. induced in the blood stream might be directly perceptible.)

Why would not the movement of the bird's wings across the earth's field induce voltages also, which would ***alternate*** with each beat of the wings? The alternating e.m.f. would give rise to small alternating charging currents that might be much more readily detectable by the bird than the rather minute effects previously discussed. Such currents, furthermore, would be directly and heavily influenced by changes in the bird's course, and could easily be the means of determining that course.

Another and somewhat analogous effect would occur when the bird, at the time of setting forth on the trip, circles aloft to "get his bearings." The e.m.f. induced by the earth's field would vary in intensity and alternate in polarity during such circling and would again create alternating charging currents. These would be smaller than the above-postulated wing currents because of the slower time-rate-of-change of voltage, but this effect would furnish a readily understandable reason for the observed fact that the birds do circle (I have not yet seen any other plausible theory), and would also be related to the vertical component of the earth's field (which seems to be receiving a lot of attention because of the observed dependence of the birds on this factor).

[1] Henry L. Yeagley, J. App. Phys. 18, 1035 (1947).

A Method for Studying the Forces between Metals and Ionic Substances

EVELYN C. MARBOE AND W. A. WEYL
The Pennsylvania State College, State College, Pennsylvania
September 22, 1948

IN connection with some work on gold ruby and related glasses a method was developed for studying the forces which exist between metal atoms and ionic substances. Today the relation between oxides or silicates and metals has become of greater and more general interest. The electrical industry is interested in combinations between oxides and metals, because the electrical properties, e.g., the temperature coefficient of electrical resistivity, of a metal can be materially changed if the metal is mixed with an electrical insulator. The surface forces of the ionic compound induce electrical fields in the metal surface and as a result the mean free path of the electrons is shortened in those areas where metal and ionic substances are in close contact. If the surface to volume ratio of the metal in contact with an ionic compound is sufficiently high, the mixture assumes semiconductor properties. Another new field involving the combination of metals with higher refractory oxides is being investigated in order to produce mechanically strong, heat resistant coatings and turbine plates.

Both fields require a better understanding of the forces acting between a metal and a solid with essentially ionic binding forces. Even if one considers the metal as a neutral, but perfectly polarizable, medium which can only reflect the forces emanating from the ionic substance (image forces), one still needs to know the magnitude of the forces emanating from the surface of the ionic substances in contact with the metal. For estimating the forces exerted between an ionic compound and a metal, the following method has been developed.

The method is based on the phenomenon that atomic gold adsorbed at the surface of a white solid has no visible light absorption. Metallic gold in relatively large crystals is yellow brown. However, when changing from the atomic subdivision to the crystalline aggregate, gold goes through a very highly colored (red or purple) stage, in which traces of this metal can be detected through its characteristic light absorption in the green part of the spectrum.

The atomic subdivision of gold is not stable at room temperature. Therefore, gold atoms adsorbed on a carrier have a tendency to form crystals through surface diffusion. The rate of diffusion depends upon the temperature and the activation energy of the process. Surface diffusion is slow when there are strong forces and fast when there are weak forces acting between the gold atoms and the surface of the carrier. This method can be used, therefore, to estimate the forces acting between gold and a carrier surface.

For this purpose the substance to be examined is treated with a dilute solution of gold chloride in water (4×10^{-3} g/ml). The gold ions are adsorbed at the surface and, by heating in hydrogen at a temperature of 150°C for a few minutes, water and HCl are removed and gold atoms are formed.

Some substances such as barium sulfate, zirconium oxide, or calcium carbonate develop a pink to purple color when heated for a certain time (approximately 1 hour) to temperatures between 200 and 800°C. The increasing aggregation of gold atoms is indicated by the formation of a faint pink color which deepens, changes into purple and gradually fades as the crystals become larger.

Other substances such as titanium dioxide or tin dioxide coated with atomic gold will not produce such a color upon heating because the forces acting between the non-noble gas-like ions, Sn^{4+} or Ti^{4+}, and the gold atoms are too strong to allow diffusion along the crystal surface. If gold is precipitated onto the surface of SnO_2 in crystals of colloidal size, further recrystallization of these crystals is prevented and a stable pigment is obtained: *Purple of Cassius*—a corresponding pigment obtained by co-precipitation of metallic gold and a substance belonging to the first group, e.g., barium sulfate, fades on heating.

Between these two extreme groups, there exist all possible transitions. Zinc oxide, for instance, exerts stronger forces upon the gold atoms than ZrO_2 or $BaSO_4$, but it still permits a slow diffusion.

This method gives a picture of the surface forces emanating from the carrier lattice. As such, it distinguishes between substances of different surface developments. It is to be expected that the forces acting between gold and a certain substance will be different for different faces of a crystal. No substance consisting of inert gas-type ions has been found which exerts strong forces upon the metal. The mere presence of non-noble gas-like ions, however, is not the only requirement for exerting strong forces towards a metal. The grouping of these ions in the lattice, their coordination number, and the nature of the anion all affect the affinity of the crystal surface towards the metal. The titanium ion in TiO_2, for example, was found to prevent diffusion nearly completely. A sodium titanium silicate, $Na_2O\times TiO_2\times SiO_2$, however, allowed the gold to aggregate.

It is not the object of this paper to discuss the nature of the binding forces between ions and metal atoms. It should be pointed out, however, that this method seems to be generally applicable. Gold is by far the most convenient metal because of its strong light absorption when present in a colloidal state. Silver, however, gives comparable results. Silver ions can be adsorbed easily on a carrier and reduced by hydrogen gas at 150°C. The silver atoms are colorless, but on some carriers they show a yellowish white fluorescence when irradiated with ultraviolet light of 3650 A.U. Upon aggregation, however, a yellow to brown color appears and the fluorescence is destroyed.

It might be interesting to mention that substances which do not exert strong forces by themselves towards metals can be made "metallophilic" by treating their surfaces with tin ions. The diffusion speed of gold on barium carbonate, for example, is relatively high. If, however, the $BaCO_3$ has been treated with a dilute solution of tin, arsenic, or antimony compounds, prior to the adsorption of the metal atoms, the diffusion is slowed down. This means that the forces between the carrier and the metal have been strengthened. It has been found empirically that treating a glass or an organic high polymer with tin compounds enhances the rate of deposition and also the adherence of a silver mirror. Other non-noble gas-like ions, such as Cu^{+} or Ti^{3+}, have similar but much weaker effects.

The Effect of "Multiple Grounds" on Electron Microscope Images

W. L. GRUBE
Research Laboratories Division, General Motors Corporation, Detroit, Michigan
October 27, 1948

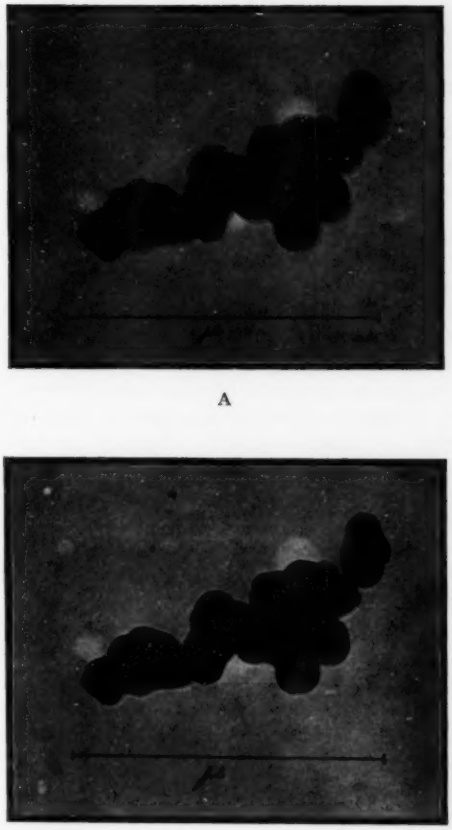
AT the recent meeting of the Electron Microscope Society of America at Toronto, Dr. James Hillier presented a paper[1] in which he discussed the various causes of image shift in electron microscopes. One of the causes of image shift, according to Dr. Hillier, is "stray magnetic fields." Whether the shift is due to such a cause can be ascertained by recording a micrograph of the same field of view with two different specimen heights. Since the "stray fields" in general affect the beam only at points below the objective lens, the shift on the plate should be the same for both specimen settings.

Although stray fields can be detected by the above method, there still remains the problem of locating the source of the fields and correcting it. It has been our experience that these fields can arise from local currents circulating throughout the microscope chassis and column.

It is general practice in electron microscope installations to ground the microscope chassis to the cooling water line as well as to some electrical neutral or conduit. When these two "grounds" are of the same potential such practice is satisfactory, but when they are not (as is generally the case) local currents are forced to circulate throughout the various metal parts of the microscope. Disturbances introduced in this manner which are alternating in character produce an image shift as shown at "A," Fig. 1.

The possibility that the disturbance affects the image indirectly through the various electrical circuits has been considered. However, there are two reasons why it is likely that the effect is direct and magnetic in nature. First, careful examination of "A," Fig. 1, shows that the image actually consists of two images which arise from a unidirectional shift of the single image shown at "B," Fig. 1. Furthermore, each image exhibits good resolution. Disturbances which act so as to upset the stability of the electrical circuits would more likely produce a general fuzziness on a single image. Second, additional column shielding sometimes improves the image quality, depending upon the path of the "ground" currents relative to the shielding. In one particular case, however, additional column shielding did not improve the image quality since the column itself was the conductor for the "ground" current.

Circulating "ground" currents can arise in many subtle ways, but in all cases they are due to the use of a "dual" ground. For example, when a separate 220-volt, 3-wire line is used to operate the forepump and the neutral is grounded to the motor frame (and in turn to the base and forepump), an all metal vacuum line to the microscope paired with the cooling water line will constitute a "dual" ground. In other cases, however, the "dual" ground arises simply from a direct connection to the microscope chassis to two possible grounds. When it is advisable, for safety reasons, to use more than one ground connection, they should be made outside the microscope and only one lead brought into the cabinet. Furthermore, the cabinet itself should be isolated from any other possible external grounds.



A

B

FIG. 1.

[1] J. Hillier, *On the Adjustment and Manipulation of the Electron Microscope*, presented at the Annual Meeting of the Electron Microscope Society of America, Toronto, Canada.

New Books

Electric Contacts

By Ragnar Holm. Pp. 398, Hugo Gebers Förlag, Stockholm, 1946.

This book is a most welcome one, for it contains the basic physical principles, as far as known, underlying the phenomena occurring in electric contacts. The subject of contacts may seem to the uninitiated to be an esoteric one, and of little concern except to those engaged in the design of relays and switching mechanisms. However, this is far from the case as a perusal of Holm's book will show. For the subject of contacts is based upon many of the principles in the fields of physics of solids, discharge in gases, and physics of surfaces. The book represents many years of careful study by one of the leading researchers in the field of contacts and is consequently thorough and authoritative. Some of the book is based upon Dr. Holm's own researches and those of his colleagues at the Research Laboratory of the Siemens Works in Berlin.

Dr. Holm has divided his book into four parts, the last one of which is a brief history of the subject. Part I deals with stationary contacts. In this part the theory and experiments relating to the resistance of a contact are carefully outlined. The idea of the constriction encountered by the flow of current through the contact is well brought out, and the important relation between electric potential and temperature in the contact constriction is developed. Other relations affecting contact resistance are also discussed, such as that between the applied contact force and resistance, the effect of corrosion on the contact surface and its breakdown through local heating in the constriction, which is referred to as coherer action. Microphonic action in carbon contacts is briefly described and rectifying contacts touched upon.

Part II, on sliding contacts, is a basic discussion of the theory and measurements of friction as a contact phenomenon. This discussion includes friction between clean metallic surfaces in vacuum, thin and thick film lubrication. The work of Bowden and his collaborators is included. Much of this section is devoted to wear in sliding contacts which is more of a practical problem than a basic physical one, but, nevertheless, a very important one to those concerned with the use of sliding contacts in commutators and other devices.

Part III deals with electrical phenomena in switching contacts and is primarily concerned with the discharges occurring between making and breaking contacts carrying low currents. Particular attention is given to arcs of short duration, and the erosion and transfer of material which they produce in contacts. The time duration of short arcs is also discussed and some circuits for quenching them are considered. The subject of bridge formation which occurs below the minimal arcing conditions is outlined, and the erosion and transfer of matter caused by bridges discussed. This last section is perhaps the most unsatisfactory part of the book, not because Dr. Holm has failed to elucidate the discharge phenomena occurring in contacts, but rather because our basic physical knowledge of discharges of short duration is inadequate. Dr. Holm has been careful to point out the many large problems yet to be solved in the field of electrical discharges relating to contact phenomena, and has made the best of our inadequate knowledge in applying it to contacts.

One of the values of this volume to the researcher will be Holm's discussion of the many unsolved problems in this field. On the other hand, many tables of data and curves are given throughout the book and in the appendix which should be of great value to the researcher and engineer alike. The book is well printed, and equations and curves are excellently displayed. The references throughout the volume are very complete and are listed in the author and literature index. The bibliography should be of great help to the researcher and engineer. It is felt that the book could have been better organized in places and much of the basic theory and experiments grouped together rather than spread out through the book. Also it would seem desirable to give more general and rigorous proofs in some of the theory; for example, the voltage-temperature relation in a contact. The reader can, of course, refer to original papers in which this is done. The reader will find typographical errors occasionally both in the equations and the writing, and at times the language may seem a bit quaint. Nevertheless, it is perfectly intelligible.

These criticisms, however, should receive small weight compared to the over-all value of the book both to the researcher and engineer. There is no other book covering this subject as completely and thoroughly as this one, except Dr. Holm's earlier book published in Germany in 1941. The present volume is a complete rewrite of the earlier German volume, and has been brought up to date. It is doubtful if another book on this subject of comparable importance will be forthcoming for many years.

Gordon Ferrie Hull, Jr.
Dartmouth College

New Booklets

Shafer Bearing Corporation, P. O. Box 57, Downers Grove, Illinois, has issued a four-page bulletin describing its self-aligning roller bearing units and featuring its new "microlock" adjustment. Free on request.

Andrew Corporation, 363 East 75th Street, Chicago 19, Illinois, has published two new bulletins. No. 85 describes the new Andrew type 1900 Automatic Dehydrator, and No. 83 announces the new Andrew line of r-f inductors. Two and four pages, free on request.

Henry A. Gardner Laboratory, Inc., 4723 Elm Street, Bethesda 14, Maryland, announces the publication of an eight-page bulletin entitled *Instruments for Measuring Appearance and Other Optical Factors*. The Laboratory does research on paint and varnish. Booklet free on request.

General Electric Company, Section 800-40, Schenectady 5, New York, has issued a four-page bulletin entitled *Pneumatic Fatigue Tester* and in it offers a free 44-page catalog on approximately 100 "widely-used specialized testing and measuring equipments." The catalog number is GEA-639.

The International Nickel Company, Inc., 67 Wall Street, New York 5, New York, has published a 36-page brochure on *Nickel Alloy Steel Castings in Industry*. Numerous photographs.

Alpha Metals, Inc., 363 Hudson Avenue, Brooklyn 1, New York, has issued its catalog No. 201 on *Alpha Tricore Solder*. Four pages, free on request.

Leeds and Northrup Company, 4934 Stenton Avenue, Philadelphia 44, Pennsylvania, has published catalog ND4B, 1948, which replaces N-OOB. The title is *L & N Pneumatic Control*. 24 pages, free on request.

The Brown Instrument Company, Wayne and Roberts Avenues, Philadelphia 44, Pennsylvania, recently issued the following three new booklets: No. 7000, *Brown Pressure and Vacuum Gauges*, 32 pages; No. 15–12, *Brown pH and Conductivity Recorders and Controllers*, 44 pages; No. 15–13, *Brown Electronik Potentiometers*, 32 pages. All free on request.

Here and There

New Appointments

Theodore C. Gams, consultant in industrial electronics, has recently been appointed Director of the School of Industrial Technology in New York City.

Edward X. Hallenberg, formerly personnel manager at Westinghouse Research Laboratories, has been named Assistant Director of the Laboratories.

Norris D. Embree is now Director of Research at Distillation Products, Inc., Rochester, New York. He had been Acting Director since June 1.

Jack W. Garrison, formerly in charge of acoustics research and development for the United States Gypsum Company, has been named assistant to the chairman of physics research at Armour Research Foundation of Illinois Institute of Technology.

Randall McG. Robertson, specialist in electronics and solid state physics, has been appointed Acting Director of the Physical Sciences Division of the Office of Naval Research. He is filling the vacancy created when Emanuel R. Piore was granted a year's leave of absence to do research at the Massachusetts Institute of Technology Research Laboratory of Electronics.

Several new appointments have been announced recently by the National Bureau of Standards. Ladislaus L. Marton is now Chief of the Electron Physics Section; Willard H. Bennett has been designated head of the Physical Electronics Section of the Atomic and Molecular Physics Division; Andrew P. Massey has been appointed head of the electronics standardization group of the Engineering Electronics Section; Harold Goldberg is the new Chief of the Ordnance Research Section, and he will be assisted by Donald Burcham; Emanuel Maxwell will do research in the Cryogenics Section of the Heat and Power Division.

Recent Awards

David S. Grey, 29, designer of lenses for television receiving sets and microscopes for medical research, received on October 22 the Adolph Lomb Medal of 1948 at the thirty-third annual meeting of the Optical Society of America. The award is given every two years to "a young scientist of outstanding promise who has made a noteworthy contribution to optics." Grey has been a member of the research staff of Polaroid Corporation, Cambridge, Massachusetts, since he was 23.

J. A. Hutcheson, Director of the Westinghouse Research Laboratories, has received the Westinghouse Order of Merit for his outstanding engineering achievements and able direction of research activities. This is the highest award of the Westinghouse Electric Corporation to its employees.

C. E. Kenneth Mees, Vice President in charge of research at Eastman Kodak Company, received November 5 the first Progress Medal of the Photographic Society of America. Dr. Mees was voted the award for his many technical, literary, and inspirational contributions to photography.

Harold F. Sherwood of Kodak Research Laboratories was chosen to receive the Rodman Medal at the 93d Annual International Exhibition of the Royal Photographic Society held recently in London. The medal is awarded for outstanding work in photomicrography, radiography, and other scientific fields. Sherwood is the first U. S. winner since the award was set up in 1935.

G. E. Building Equipment for Indian Institute

An a.c. network analyzer, for installation at the Indian Institute of Science, Bangalore, India, is under construction at the General Electric Company's Philadelphia Works. When completed early next year the instrument will be installed in a newly built, high voltage laboratory which is expected by Indian engineers to be one of the most modern in Asia, and which is part of India's engineering training program.

American Mathematical Society

A meeting of the American Mathematical Society was held at New York University on October 30. There were sessions for contributed papers, both in pure and applied mathematics, covering the fields of analysis, algebra, topology, and logic. The principal address, "Convexity Theorems," was given by Professor Raphael Salem of the Massachusetts Institute of Technology.

Seminars on High Polymers

The National Bureau of Standards is continuing this winter its three-year custom of sponsoring seminars on high polymers. Dr. Robert Simha of the Division of Organic and

Fibrous Materials is chairman of the program. Remaining lectures in the series are as follows:

February 24: "Some aspects of dynamic rubber-like elasticity," A. W. NOLLE, *University of Texas.*

March 3: "Reactions of free radicals with hydrocarbons," E. W. R. STEACIE, *National Research Council.*

April 7: "Variables which influence the properties of chemical rubbers prepared by emulsion polymerization," C. F. FRYLING, *Phillips Petroleum Company.*

May 5: "The chemistry of some derived polymers of the vinyl series," W. O. KENYON, *Eastman Kodak Company.*

Predoctoral Fellowship for Women

Sigma Delta Epsilon, graduate women's scientific fraternity, is preparing to award its third fellowship to a woman student in mathematics, physical science, or biological science. Application blanks for the $1200 award may be obtained from Dr. Virginia Bartow, 7 Chemistry Annex, University of Illinois, Urbana, Illinois. Applications for the year 1949–50 should be submitted before February 1, 1949.

Sound Apparatus Company Moves

Sound Apparatus Company of New York City has consolidated its main office and manufacturing plant in Stirling, New Jersey. Its developmental laboratory at Millington, New Jersey, nearby, is being enlarged. It is equipped with facilities for consulting and experimenting services in the field of acoustical and electrical measurements.

Committee on Plans for Mobilizing Science Formed

A Committee on Plans for Mobilizing Science is studying the problem of organization of scientific resources in event of a national emergency, Dr. Vannevar Bush, retiring Chairman of the Research and Development Board, National Military Establishment, announced in mid-October. The newly formed committee, which is headed by Dr. Irvin Stewart, President of West Virginia University, was asked by Dr. Bush to make recommendations on the role the Research and Development Board would play under wartime conditions and to suggest plans for the integration of the present organization with such additional organization as may seem desirable in time of war. Attending the first meeting of the committee on October 13 and 14 were the chairman, Dr. Stewart, Dr. James P. Baxter, III, President of Williams College, Mr. John T. Connor, Secretary and Counsel for Merck and Company, Inc., Rahway, New Jersey, and Dr. Willard Machle of New York University.

Research Corporation Expanding Midwestern Services

Research Corporation is expanding its midwestern office in Chicago, to facilitate grants-in-aid for the advancement of science in midwestern educational and scientific institutions. Walter E. Thwaite, Jr., field representative for patent management and grants-in-aid, will be in charge of the new phases of operation in the Chicago office. Harry M. Pier is manager of the office.

Symposium on Luminescence

The Optical Society of America is planning a Symposium on Luminescence as a feature of its Winter Meeting next March. The meeting will be held in New York City on March 10–12, 1949 and one day of the general meeting will be devoted to the symposium. Those who are working on luminescence are invited to avail themselves of the opportunity of partaking in it, whether members of the Society or not. Plans for the symposium will be facilitated if those desiring to contribute will communicate as early as possible with Dr. Gorton R. Fonda, Research Laboratory, General Electric Company, Schenectady, New York.